VOLUME 102 PART B NUMBER 5 SEPTEMBER 1955

The Proceedings

OF

THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871: INCORPORATED BY ROYAL CHARTER 1921

PART B

RADIO AND ELECTRONIC ENGINEERING
(INCLUDING COMMUNICATION ENGINEERING)

SAVOY PLACE . LONDON W.C.2

Price Seven Shillings and Sixpence

The Institution of Electrical Engineers

FOUNDED 1871
INCORPORATED BY ROYAL CHARTER 1921

PATRON: HER MAJESTY THE QUEEN

COUNCIL 1954-55

President

J. ECCLES, C.B.E., B.Sc.

Past-Presidents

SIR JAMES SWINBURNE, BART., F.R.S.
W. H. ECCLES, D.Sc., F.R.S.
THE RT. HON. THE EARL OF MOUNT EDGCUMBE, T.D.
J. M. DONALDSON, M.C.
PROF. E. W. MARCHANT, D.Sc.
P. V. HUNTER, C.B.E.
H. T. YOUNG.
SIR GEORGE LEE, O.B.E., M.C.
SIR ARTHUR P. M. FLEMING, C.B.E., D.Eng., LL.D.
J. R. BEARD, C.B.E., M.Sc.
SIR NOEL ASHBRIDGE, B.Sc.(Eng.).
COLONEL SIR A. STANLEY ANGWIN, K.B.E., D.S.O., M.C., T.D., D.Sc.(Eng.).
SIR HARRY RAILING, D.Eng.
P. DUNSHEATH, C.B.E., M.A., D.Sc.(Eng.).
SIR VINCENT Z. DE FERRANTI, M.C.
T. G. N. HALDANE, M.A.
PROF. E. B. MOULLIN, M.A., Sc.D.
SIR ARCHIBALD J. GILL, B.Sc.(Eng.).
SIR JOHN HACKING.
COLONEL B. H. LEESON, C.B.E., T.D.
SIR HAROLD BISHOP, C.B.E., B.Sc.(Eng.).

Vice-Presidents

T. E. GOLDUP, C.B.E.
S. E. GOODALL, M.Sc.(Eng.).
WILLIS JACKSON, D.Sc., D.Phil., F.R.S.
SIR GEORGE H. NELSON, BART.
SIR W. GORDON RADLEY, C.B.E., Ph.D.(Eng.).

Honorary Treasurer

H. W. GRIMMITT.

Ordinary Members of Council

J. BENNETT.
A. R. COOPER.
A. T. CRAWFORD, B.Sc.
C. DANNATT, O.B.E., D.Sc.
B. DONKIN, B.A.
O. W. HUMPHREYS, B.Sc.
C. R. KING, C.B.E.
H. R. L. LAMONT, Ph.D., M.A., B.Sc.
F. J. LANE, O.B.E., M.Sc.
G. S. C. LUCAS, O.B.E.
G. LYON, M.Sc.(Eng.).
SIR HAMISH D. MACLAREN, K.B.E., C.B., D.F.C., LL.D., B.Sc.
A. H. MUMFORD, O.B.E., B.Sc.(Eng.).
W. F. PARKER.
PROF. M. G. SAY, Ph.D., M.Sc.
R. L. SMITH-ROSE, C.B.E., D.Sc., Ph.D.
G. O. WATSON.
J. WESTCOTT, B.Sc.(Eng.), Ph.D.
E. L. E. WHEATCROFT, M.A.
R. T. B. WYNN, C.B.E., M.A.

Chairmen and Past-Chairmen of Sections

Measurements:
M. WHITEHEAD.
*J. F. COALES, O.B.E., M.A.

Radio:
C. W. OATLEY, M.A., M.Sc.
*J. A. SMALE, C.B.E., A.F.C., B.Sc.

Supply:
J. D. PEATTIE, B.Sc.
*L. G. BRAZIER, Ph.D., B.Sc.

Utilization:
J. I. BERNARD, B.Sc.Tech.
*B. L. METCALF, B.Sc.(Eng.).

Chairmen and Past-Chairmen of Local Centres

East Midland Centre:
J. H. MITCHELL, B.Sc., Ph.D.
*C. D. WILKINSON.

North Midland Centre:
W. A. CROCKER.
*G. CATON.

North-Western Centre:
PROF. E. BRADSHAW, M.B.E., M.Sc.Tech., Ph.D.
*H. WEST.

Scottish Centre:
J. S. HASTIE, B.Sc.(Eng.).
*C. H. A. COLLYNS.

Mersey and North Wales Centre:
P. R. DUNN, B.Sc.
*T. COATES, M.Eng.

North-Eastern Centre:
G. W. B. MITCHELL, B.A.
*H. ESTHER, B.Eng.

Northern Ireland Centre:
MAJOR P. L. BARKER, B.Sc.
*J. R. W. MURLAND, B.Sc.(Eng.).

South Midland Centre:
A. R. BLANDFORD.
*H. J. GIBSON, B.Sc.

Southern Centre:
E. A. LOGAN, M.Sc.
*COMDR. C. V. ROBINSON, R.N., O.B.E.

Western Centre:
A. N. IRENS.
*J. VAUGHAN HARRIES.

* Past-Chairman.

RADIO SECTION COMMITTEE 1954-55

Chairman

C. W. OATLEY, M.A., M.Sc.

Vice-Chairmen

H. STANESBY.
R. C. G. WILLIAMS, Ph.D., B.Sc.(Eng.).

Past-Chairmen

J. A. SMALE, C.B.E., A.F.C., B.Sc.
E. C. S. MEGAW, M.B.E., D.Sc.

Ordinary Members of Committee

PROF. H. E. M. BARLOW, Ph.D., B.Sc.(Eng.).
N. R. BLIGH, B.Sc.(Eng.).
G. G. MACFARLANE, Dr.Ing., B.Sc.
B. N. MACLARTY, O.B.E.
J. H. H. MERRIMAN, M.Sc.
BRIG. E. J. H. MOPPETT.
H. PAGE, M.Sc.
A. R. A. RENDALL, Ph.D., B.Sc.
N. C. ROBERTSON, C.M.G., M.B.E.
L. RUSHFORTH, M.B.E., B.Sc.
J. A. SAXTON, D.Sc., Ph.D.
A. M. THORNTON, B.Sc.

And

The President (*ex officio*).
The Chairman of the Papers Committee.
A. H. MUMFORD, O.B.E., B.Sc.(Eng.) (representing the Council).
G. E. MIDDLETON, M.A. (representing the Cambridge Radio Group).
J. MOIR (representing the South Midland Radio Group).
D. H. THOMAS, M.Sc.Tech., B.Sc.(Eng.) (representing the North-Eastern Radio and Measurements Group).

The following nominees of Government Departments:
Admiralty: CAPTAIN G. C. F. WHITAKER, R.N.
Air Ministry: AIR COMMODORE G. H. RANDLE, B.A.
Department of Scientific and Industrial Research: A. F. WILKINS, O.B.E., M.Sc.Tech.
Ministry of Supply: F. S. BARTON, C.B.E., M.A., B.Sc.
Post Office: CAPTAIN C. F. BOOTH, O.B.E.
War Office: COL. J. H. E. PANTON, O.B.E.

Secretary

W. K. BRASHER, C.B.E., M.A., M.I.E.E.

Deputy Secretary

F. JERVIS SMITH, M.I.E.E.

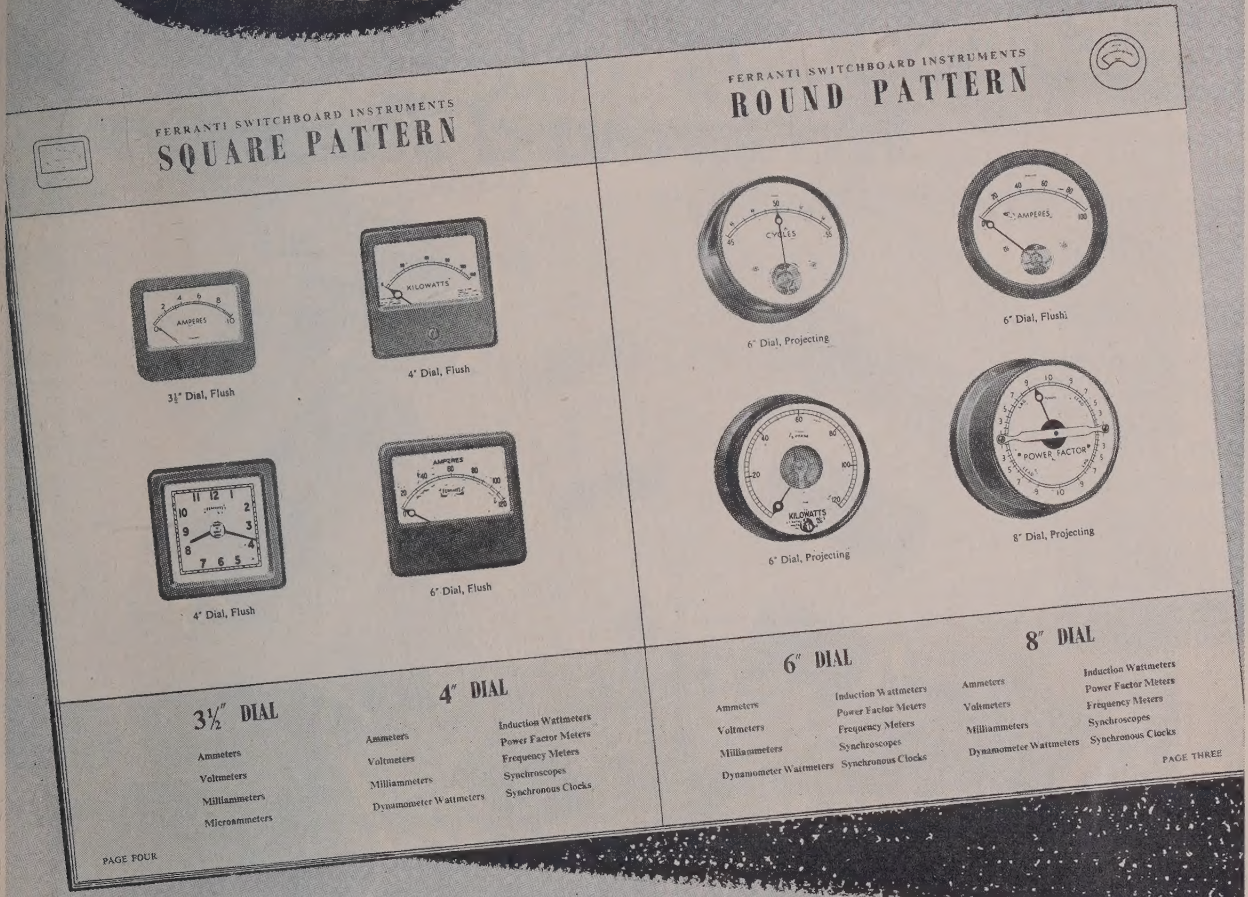
Assistant Secretary

F. C. HARRIS.

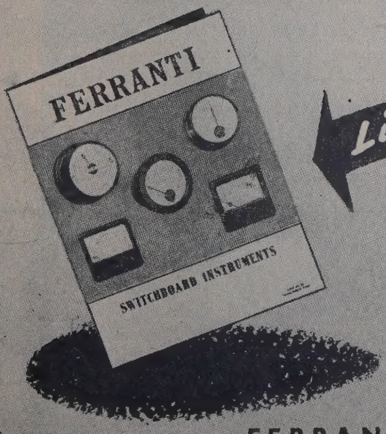
Editor-in-Chief

G. E. WILLIAMS, B.Sc.(Eng.), M.I.E.E.

FERRANTI *Switchboard Instruments*



Illustrated above are several Ferranti Switchboard Instruments from a wide range. Shallow pressed steel cases $3\frac{1}{4}''$, $3\frac{1}{2}''$, $4''$, $6''$ and $8''$ round or square dial. Projecting or flush mounting. All types of Ferranti switchboard instruments are also available in hermetically sealed cases.



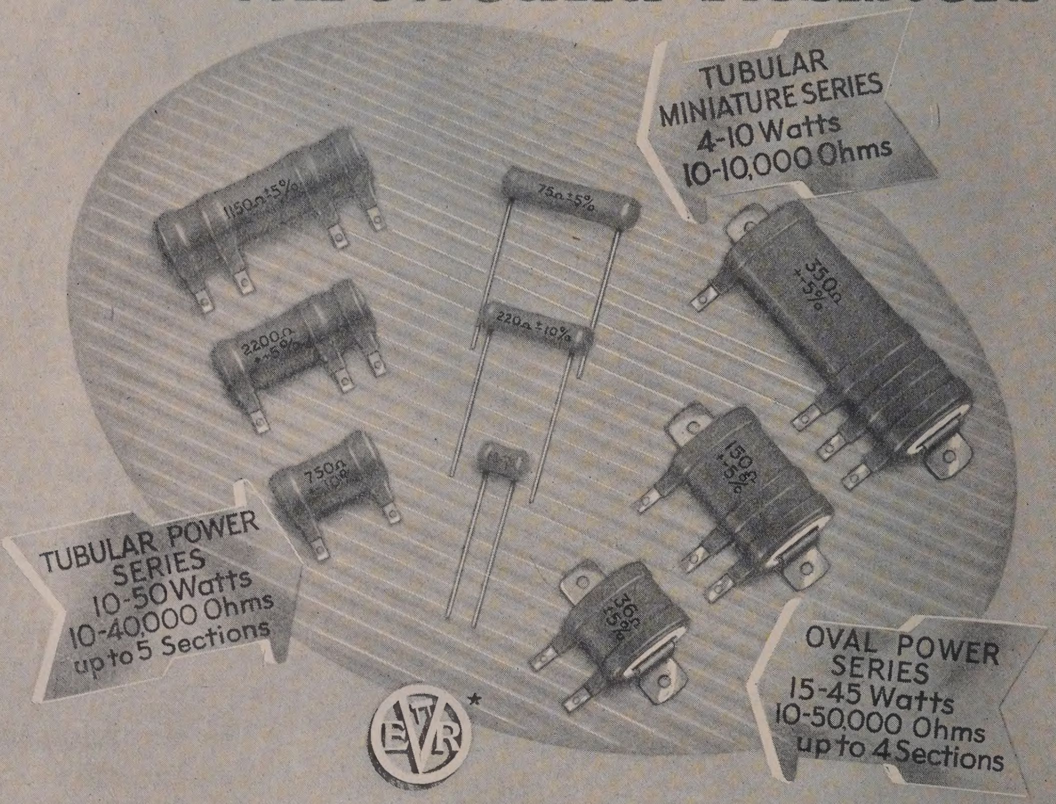
List IN30 and other catalogues supplied on request

FERRANTI

FERRANTI

FERRANTI LTD · MOSTON · MANCHESTER 10
Head Office and Works: HOLLINWOOD · LANCS
London Office: KERN HOUSE, 36 KINGSWAY, W.C.2.

ERIE* "Silertex" Wirewound Resistors



The excellence of Erie wirewound resistors is due, not only to the sound workmanship always associated with Erie products, but also to the coating of silicone bonded phenolic resin, known as "Silertex", which has all the advantages of vitreous enamel, plus the added advantage of a much lower firing temperature, and consequent freedom from surface crazing and, above all, from oxidation of the resistance wire.

That is why they are reliable above all others, and that is why you will find them used by those manufacturers who take pride in quality.

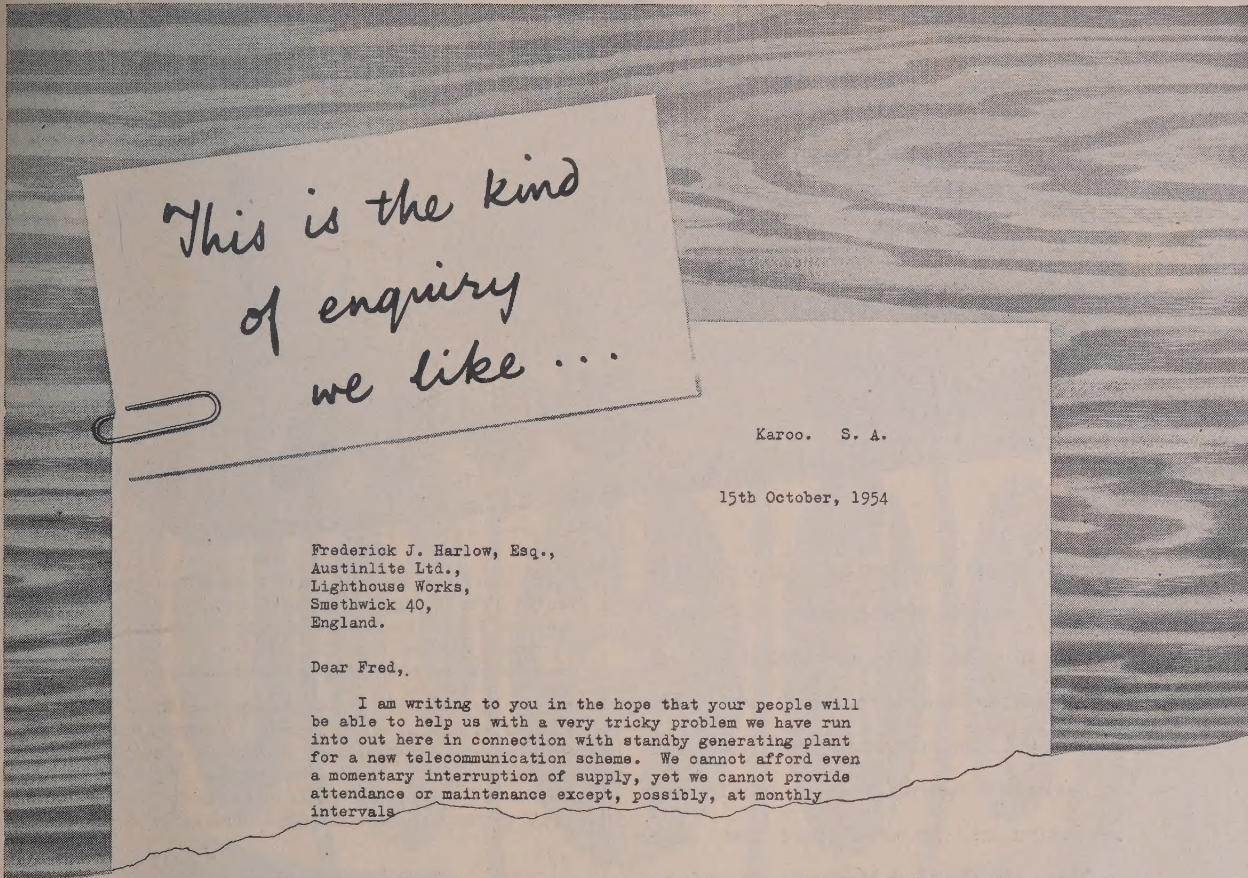
SERVICE COSTS MONEY – SO FIT "SILERTEX" TODAY

ERIE* Resistor Ltd

*Registered Trade Marks

Carlisle Road, The Hyde, London, N.W. 9., England. Telephone: COLindale 8011.
Factories: London and Great Yarmouth, England; Trenton, Ontario, Canada; Erie, Pa., and Holly Springs, Miss., U. S. A.

Automatic Generating Plant



The more tricky the problem, the more arduous the conditions under which the equipment must operate, the better we like it. Difficult, unusual generating plant is our metier. It does not matter how long it must run without attention. It does not matter whether there is a mains supply or not. Nor if the supply is erratic and unreliable. Even if the requirements do not come within our standard range Austinlite plant can be designed which will ensure continuous, steady and extremely reliable supply.

During the past 25 years Austinlite plant has been steadily developed, always with the emphasis on quality and dependability. We can now provide automatic generating plant of many types in sizes from 1.4 to 250kVA, for continuous unattended operation, or for mains standby work — to the strictest no break specification where necessary. However difficult a particular problem may be, we have good reason for the confidence with which we affirm that Austinlite can tackle it.

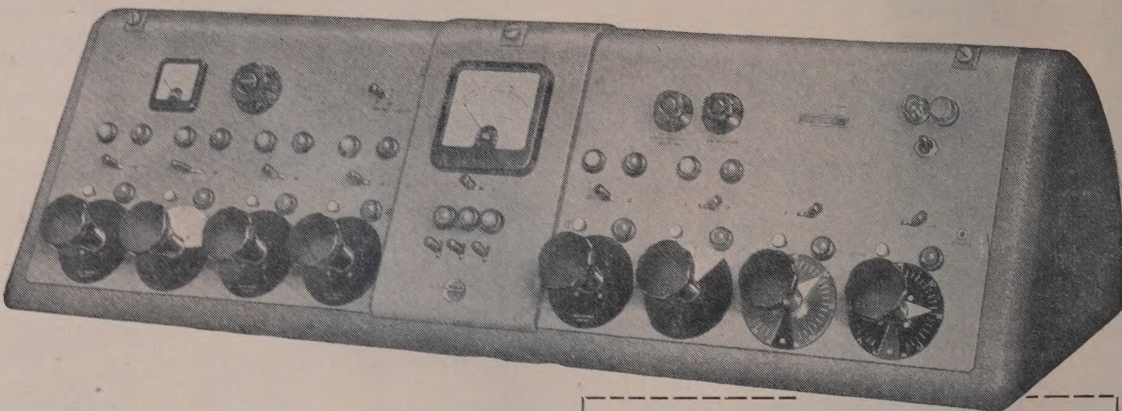
In this country our plant is used by the G.P.O., the Ministry of Civil Aviation and British Railways. It is installed in many parts of the world in lighthouses where reliability is vital. It is in operation in South Africa, Australia, Denmark, the U.S.S.R., Malta, Syria, New Zealand and Norway; as well as a number of tropical countries including Nigeria, Malaya and Borneo.

Austinlite AUTOMATIC GENERATING PLANT Tailor-made by STONE-CHANCE LTD.

(The makers of Sumo Pumps and Stone-Chance Lighthouses)

LIGHTHOUSE WORKS, SMETHWICK 40, BIRMINGHAM

Marconi Complete SOUND STUDIO EQUIPMENT

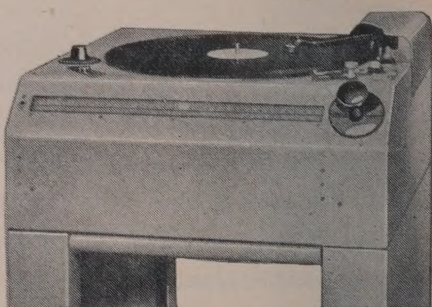


Marconi sound broadcasting equipment provides an extensive, flexible and versatile range of units from cue lamps and control consoles to automatic monitors and aerials. It covers every phase of the common operating requirements of the majority of systems. No two broadcasting administrations, however, have the same problems and therefore Marconi's are ready to engineer particular schemes and can meet every requirement in AM and FM broadcasting.

The first advertised sound broadcast was made from the Marconi transmitter at Chelmsford in June 1920. Today 75% of the countries in the world rely on Marconi broadcasting equipment.

Shown above is the Control Console Type BD 501 which handles input of programme material from two studios, announcer's microphone, a local microphone, several disc reproducers and four O.B. lines. Two output channels for rehearsal and programme conditions are provided.

Below is the heavy duty Disc Reproducer Type BD503B, designed for accurate groove location. A 3-speed turntable and long arm lightweight pick-up are employed, and a universal corrector unit is incorporated. The novel features of this equipment are the optical groove locator and an automatic raise/lower device coupled to the fader.



Lifeline of communication

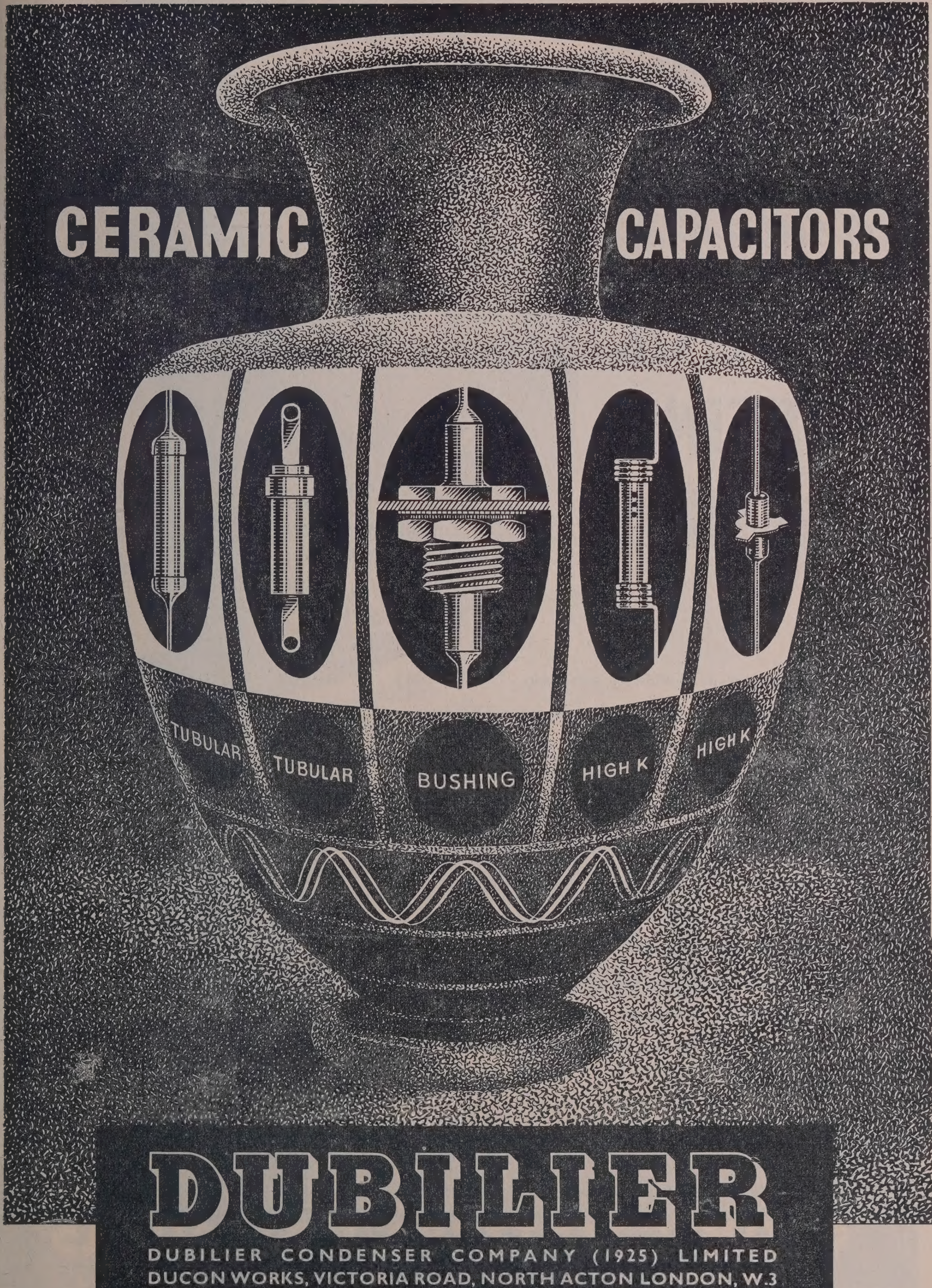
MARCONI

Complete Broadcasting and Television Systems

MARCONI'S WIRELESS TELEGRAPH CO. LTD., CHELMSFORD, ESSEX

Partners in progress with The 'ENGLISH ELECTRIC' *Company Limited*

LB3


CERAMIC
CAPACITORS
TUBULAR
TUBULAR
BUSHING
HIGH K
HIGH K
DUBILIER
DUBILIER CONDENSER COMPANY (1925) LIMITED
DUCON WORKS, VICTORIA ROAD, NORTH ACTON LONDON, W.3
Telephone: ACOrn 2241 Telegrams: Hivoltcon Wesphone London

DN. 134

good magnetic characteristics

demand

CAREFUL CASTING CONTROL

Standard **maintains its established leadership in the manufacture of high permeability magnetic alloys by constant vigilance in the control of each and every production process, one of which is illustrated here. Produced by a Company which has the unique advantage of being a large-scale user of its own magnetic materials, a long experience of the applications of these materials gives full appreciation of the properties essential for uniform electrical characteristics and stable performance.**

It will pay you to investigate the capabilities of *Standard* magnetic alloys with relation to your specific requirements.

- PERMALLOY 'C' for highest initial permeability, useful for wide-band frequency transformers, current transformers, chokes, relays and magnetic shielding.
- PERMALLOY 'B' has lower initial permeability than Permalloy 'C' but higher values of flux density. Suitable where high permeability to alternating field is required superimposed upon a steady polarising field.
- PERMALLOY 'D' for very high resistivity without undue lowering of the maximum flux density. Variation of permeability with frequency is small. Ideal for H.F. applications.
- PERMALLOY 'F' for high flux density, very rectangular hysteresis loop, with a retentivity of at least 95% of its saturation value and low coercive force. Ideal for saturable reactors, magnetic amplifiers, digital computors, memory devices, etc.
- V-PERMENDUR for high permeability with a very high value of maximum flux density. Finds special application for use as high quality receiver diaphragms, also motor generators and servo-mechanisms in aircraft where weight and volume are important factors.

Standard Telephones and Cables Limited

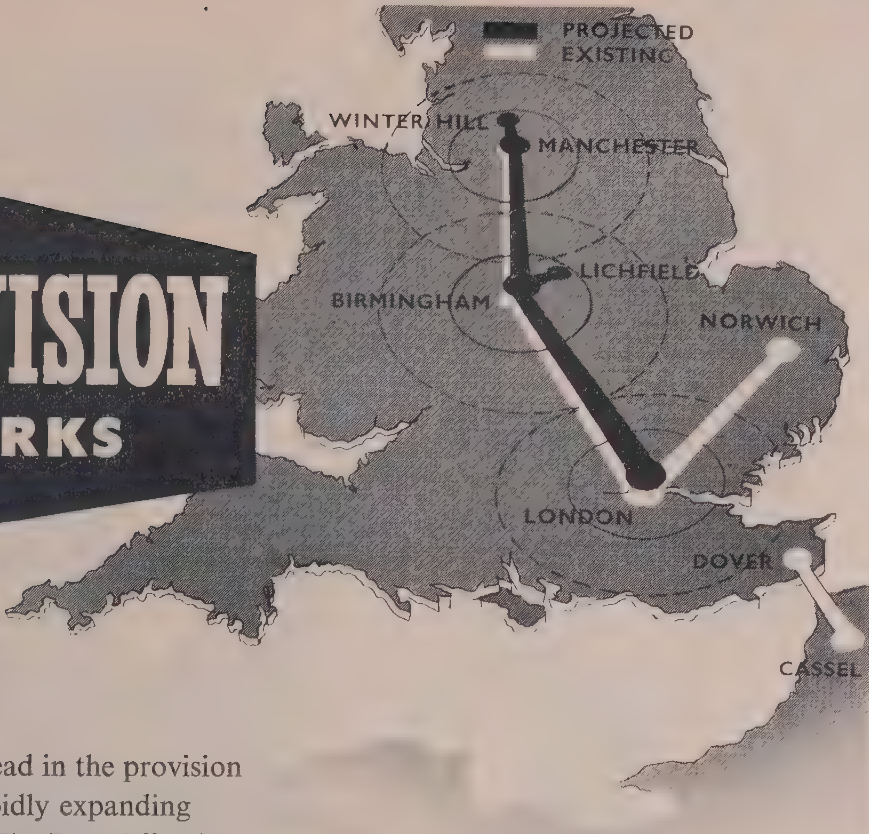
Registered Office: Connaught House, Aldwych, W.C.2

TELEPHONE LINE DIVISION : North Woolwich, London, E.16


CATHODEON
Quartz Crystals
FOR RELIABLE FREQUENCY CONTROL
Frequency Range 2,000—20,000 kc/s
Our range now includes crystals for close tolerance requirements
Enquiries are invited for overtones up to 60 Mc/s
CATHODEON CRYSTALS LIMITED
LINTON · CAMBRIDGESHIRE · Telephone LINTON 223


G.E.C.
EQUIPMENT

for TELEVISION NETWORKS



G.E.C. continue to lead in the provision of links for Britain's rapidly expanding T.V. networks. The Post Office is now being supplied with equipment for additional coverage in the Midlands and North-Western Areas against the future requirements of the Independent Television Authority. (G.E.C. equipment already conveys B.B.C. television programmes to these areas). The new coverage being provided, by radio relay or by coaxial cable, is as follows:—

THE MIDLAND LINK

The existing G.E.C. microwave link—the first in Britain, installed in 1946—between London and Birmingham, is the basis of the new network. A spur to Lichfield, using the G.E.C. SPO 5551 radio relay system, is being added.

THE NORTH-WESTERN LINK

The network is being extended to the North-West by the addition of G.E.C. terminal translating and coaxial line equipment to serve the I.T.A. North-Western Area transmitter, over cable from Birmingham to Manchester and Winter Hill. This route includes 4 terminal and 18 repeater stations, using wideband amplifier equipment with supervision at the Manchester station.

The SPO 5551 two-channel transmit terminal equipment in Telephone House, Birmingham (London-Lichfield link) during commissioning tests.

G.E.C. OF ENGLAND

THE GENERAL ELECTRIC CO LTD OF ENGLAND
TELEPHONE WORKS COVENTRY

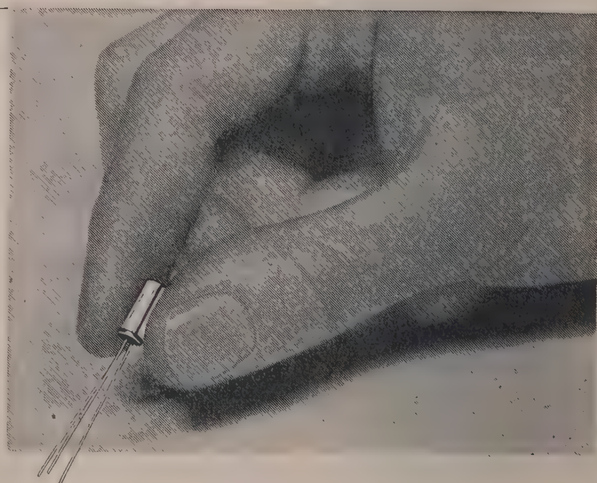
Recent developments in G.E.C. Semi-Conductors

Three new G.E.C. Junction Transistors

Now available to Electronic Equipment Manufacturers

EW53	EW58	EW59

The performance of junction transistors at high frequencies is partly limited by the reduction of current gain with frequency. A more important frequency limitation is often set by the high values of collector capacitance c_c and "extrinsic" base resistance r_{bo}. These new transistors have been designed to have a particularly low value of r_{bo} with consequent improvement in high frequency performance.



EXAMPLE: At a frequency of 465 kc/s, operating between a 50 Ω source and a 10 kΩ load, using an EW59 transistor, a power gain of the order of 20db is possible.

Added ruggedness combined with freedom from moisture is achieved by complete enclosure in gold-plated metal envelopes.

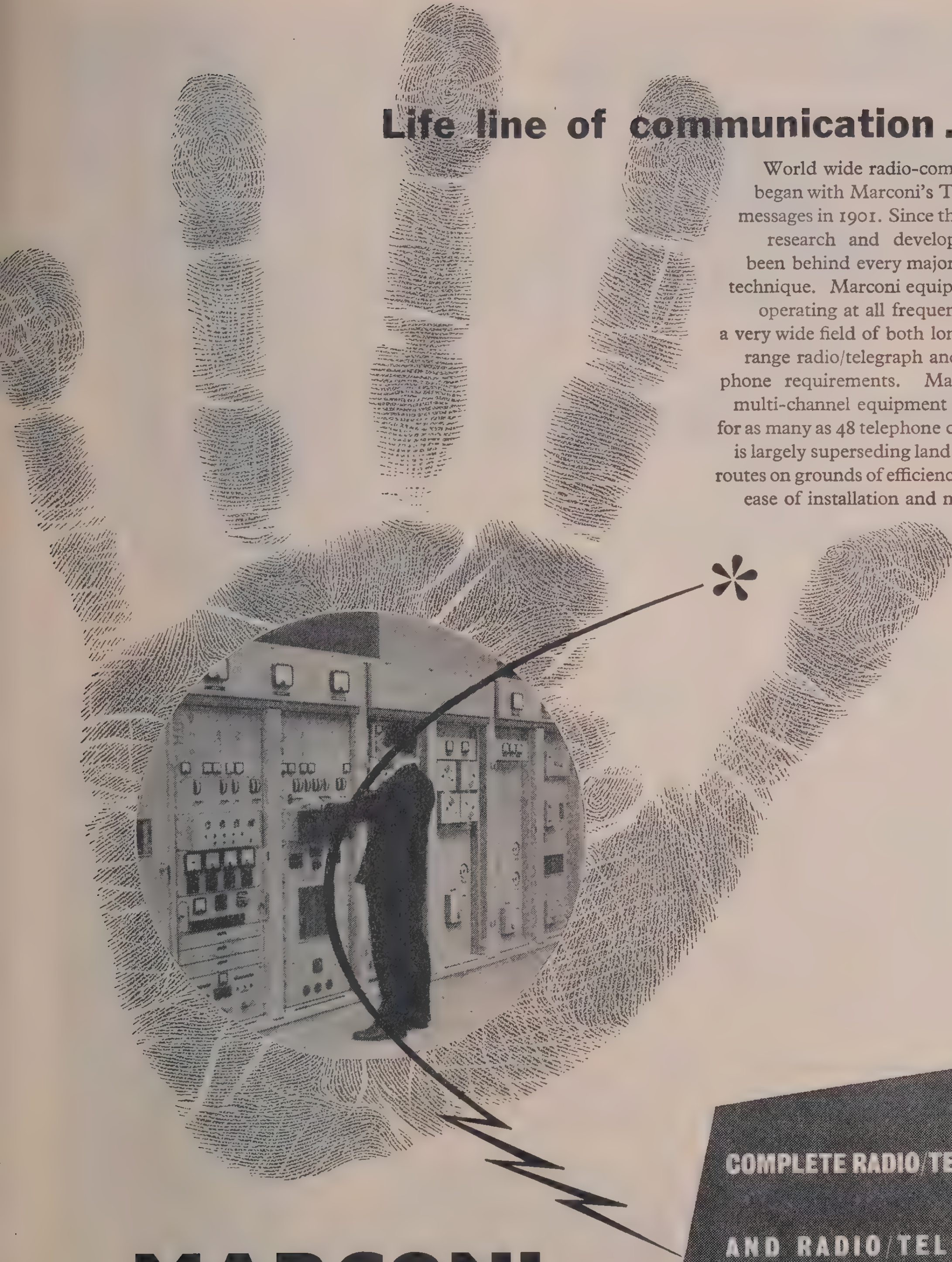
Other recent G.E.C. semi-conductors are the EW51 H.F. Point Contact Transistor and the EW54 medium power Junction Rectifier.

You are invited to write to
THE OSRAM VALVE & ELECTRONICS DEPARTMENT
about your transistor application problems.

THE GENERAL ELECTRIC CO. LTD., MAGNET HOUSE, KINGSWAY, LONDON, W.C.2

Life line of communication . . .

World wide radio-communication began with Marconi's Transatlantic messages in 1901. Since then Marconi research and development have been behind every major advance in technique. Marconi equipment today, operating at all frequencies, covers a very wide field of both long and short range radio/telegraph and radio/telephone requirements. Marconi VHF multi-channel equipment can provide for as many as 48 telephone channels and is largely superseding land line or cable routes on grounds of efficiency, economy, ease of installation and maintenance.



COMPLETE RADIO/TELEPHONE AND RADIO/TELEGRAPH SYSTEMS AND EQUIPMENT

MARCONI

COMPLETE COMMUNICATION SYSTEMS

Surveyed, Planned, Installed, Maintained

MARCONI'S WIRELESS TELEGRAPH COMPANY LIMITED · CHELMSFORD · ESSEX

LG I

Marconi Surveying Service

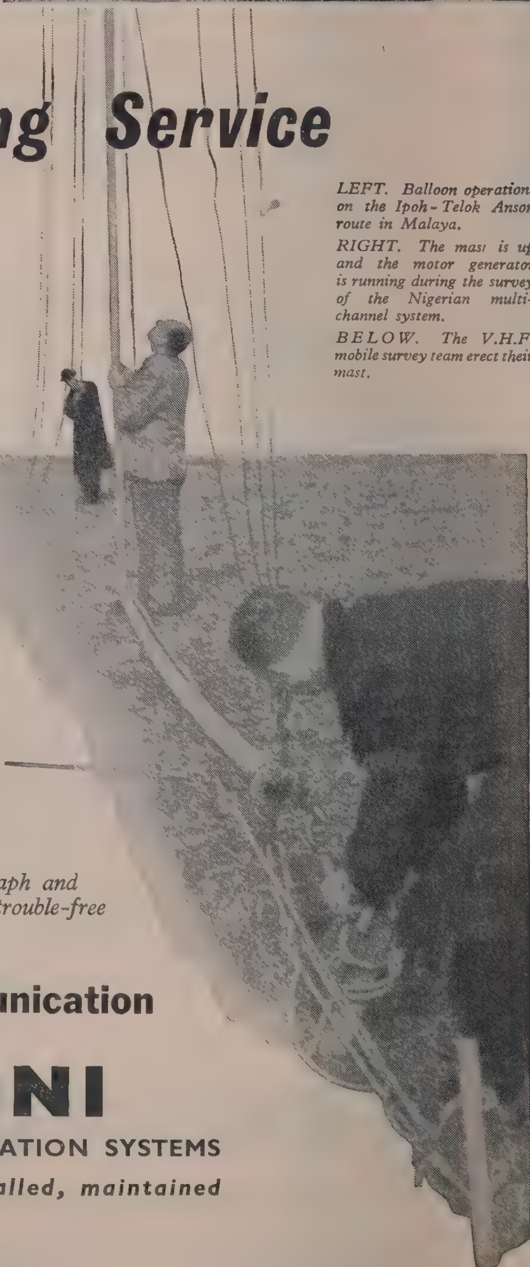
LEFT. Balloon operations on the Ipoh-Telok Anson route in Malaya.

RIGHT. The mast is up and the motor generator is running during the survey of the Nigerian multichannel system.

BELOW. The V.H.F. mobile survey team erect their mast.

Before planning any communication system, and particularly a microwave or V.H.F. multichannel system, a survey of the propagation conditions over the proposed path or area is essential. Similar, but less exhaustive surveys, are also necessary before planning V.H.F. mobile systems. Such surveys are undertaken by Marconi's, one of the very few radio manufacturers who do so. The teams engaged in the work may be called upon to operate in desert, swamp and jungle, over which line and cable routes would be impractical, on windswept moorlands or in densely populated city and suburban areas. Surveys are being, or have already been carried out all over the world, including: Uganda, Kenya, Tanganyika, Nigeria, Gold Coast, Tangier, Azores Norway, Turkey, Greece, Malaya, Ceylon, West Indies, Sweden, and also, of course, in Britain.

Over 80 countries now have Marconi-equipped telegraph and communications services. Many of these are still giving trouble-free service after more than twenty years in operation.



Lifeline of communication

MARCONI

COMPLETE COMMUNICATION SYSTEMS

Surveyed, planned, installed, maintained

MARCONI'S WIRELESS TELEGRAPH CO., LTD., CHELMSFORD, ESSEX

LC 10

TYPE HU II

Marconi VF operated Telegraph Recording Unit

The HU unit shown here is for bench mounting. It can also be built in to a standard rack or cabinet containing the associated receiver. It is designed in accordance with the Marconi practice of providing integrated units which can be rearranged when an initial basic installation is expanded to provide additional facilities

The Type HU II recording unit (Frequency Shift Adapter) is intended for use in association with a suitable communication receiver of on/off or frequency shift transmissions.

The AF output of the associated receiver is converted into double-current signals suitable for operating an undulator, relay or similar equipment requiring a space-mark current up to 30 mA. The only modification necessary in the receiver is readjustment of the BFO frequency in some cases.

A double-current teleprinter may be operated directly and single-current teletype machines through a polarised relay.

Over 80 countries now have Marconi equipped telegraph and communication systems. Many of these are still giving trouble free service after more than 20 years in operation.

Lifeline of communication

MARCONI

COMPLETE COMMUNICATION SYSTEMS

Surveyed, planned, installed, maintained

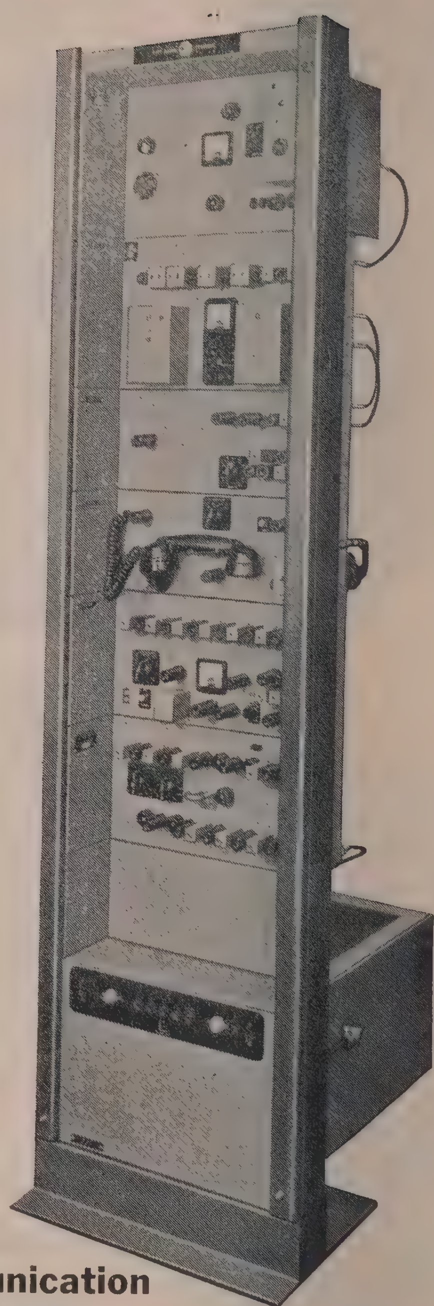
MARCONI'S WIRELESS TELEGRAPH CO. LTD., CHELMSFORD, ESSEX

Partners in progress with The **'ENGLISH ELECTRIC'** *Company Limited*

LC II

Marconi VHF Multi-Channel Equipment

TYPE HM 181



Multi-channel radio links are not only recognised economic alternatives to line and cable routes wherever the latter are costly because of intensive urban development or the wild nature of the terrain; they are frequently preferable in their own right. The type HM 181 equipment has been designed for comparatively simple schemes using two terminals working point-to-point or with a limited number of repeaters. It operates in the frequency range 150-200 Mc/s, employs frequency modulation and gives high performance with low distortion.

It provides the following facilities:—

- 8, 16 or 24 channels
- Repeaters with easy channel dropping facilities
- Unattended operation
- Engineers' order wire
- Ease of access for maintenance

Over 80 countries now have Marconi equipped telegraph and communication systems. Many of these are still giving trouble free service after more than twenty years in operation

Lifeline of communication

MARCONI

COMPLETE COMMUNICATION SYSTEMS

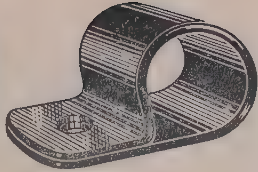
Surveyed, planned, installed, maintained

MARCONI'S WIRELESS TELEGRAPH CO., LTD., CHELMSFORD, ESSEX

Partners in progress with The **'ENGLISH ELECTRIC'** *Company Ltd.*

LC 7

PLASKLIP



Cable Clip . . .

. . Gives complete security

This non-metallic high-dielectric cable clip provides the safe means of securing cable looms and components in all radio and electrical equipment. The Plasklip is manufactured in a very extensive range covering all wiring requirements. Made in non-magnetic materials with radiused edges. Fully tropical. Approved all services.

the quick Fix BUSH

Saves time and money

Non-metallic · Fully Tropical

Here is a Bush designed for instant assembly by a simple snap on finger action. Completely secure under all working conditions. High dielectric. Approved all services.

Samples and literature available on request.

INSULOID MANUFACTURING COMPANY LIMITED

Sharston Works, Leaston Ave, Wythenshawe, Manchester. Tel.: Wythenshawe 2842

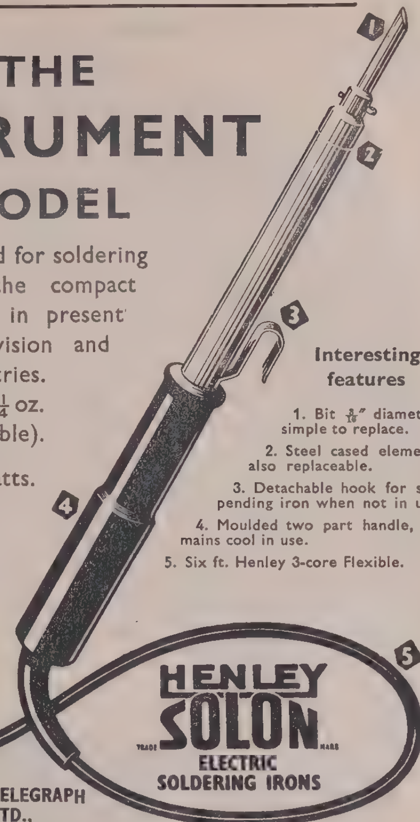
THE INSTRUMENT MODEL

Specially designed for soldering operations in the compact assemblies used in present day radio, television and electronic industries.

Weight $3\frac{1}{4}$ oz. (excluding flexible).

Length 9 in. 25 Watts.

Voltage Range
12, 24, 50, 100/110, 120/130, 200/220 and 220/240.



Interesting features

1. Bit $\frac{3}{16}''$ diameter, simple to replace.
2. Steel cased element, also replaceable.
3. Detachable hook for suspending iron when not in use.
4. Moulded two part handle, remains cool in use.
5. Six ft. Henley 3-core Flexible.

W. T. HENLEY'S TELEGRAPH WORKS CO., LTD.,

51/53, Hatton Garden, London, E.C.1

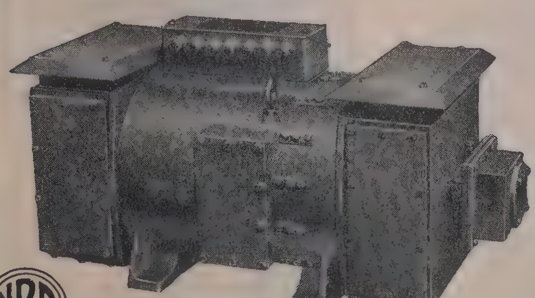
NEWTON-DERBY

ELECTRICAL EQUIPMENT

High Frequency Alternators

(Send for Publication No. 1003/2)

Also makers of Rotary Transformers and Anode Converters, Wind and Engine-Driven Aircraft Generators, High Tension D.C. Generators, and Automatic Carbon Pile Voltage Regulators.



NBD

$2\frac{1}{2}$kVA Motor Alternator. Drip proof to 45°. Motor 220 volts D.C. Output 120 volts. 3 phase. 333 cycles per second. Motor includes an automatic constant speed governor. Weight 450 lb.

NEWTON BROTHERS (DERBY) LTD

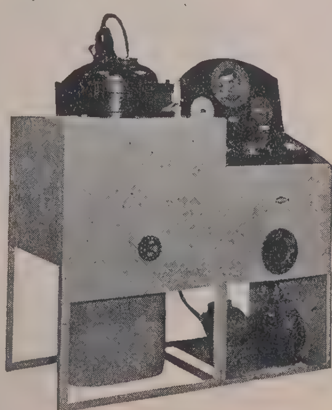
HEAD OFFICE & WORKS: ALFRETON ROAD, DERBY
TELEPHONE: DERBY 47676 (4 lines) TELEGRAMS: DYNAMO, DERBY
LONDON OFFICE: IMPERIAL BUILDINGS, 56 KINGSWAY W.C.2

IMPREGNATE

your coils with ease

BLICKVAC

High Vacuum Impregnators meet the most stringent specifications and yet are easy to handle. Full range of models available to meet the needs of the large-scale producer, the research laboratory or the small Rewind shop.



Outstanding Features:

- Ease in control
- Ease in cleaning
- Elimination of vibration
- Unequalled flexibility and performance
- Simple attachment of auxiliary autoclaves
- Units available suitable for Varnish, Wax, Bitumen and Potting Resins. Users include M.O.S., N.C.B., G.E.C., Pye, Marconi, Metro-Vick.

If your problem is COIL IMPREGNATION or impregnating or casting with Potting Resins consult:

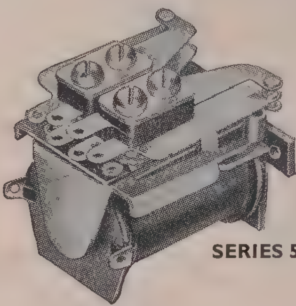
BLICKVAC ENGINEERING LTD

Bede Trading Estate, Jarrow, Co. Durham — Jarrow 89–7155
96–100 Aldersgate Street, London, E.C.1. — Monarch 6256/8

RELAYS

HERMETICALLY SEALED

"Series 595H" is a hermetically sealed version of our series 595 Relay, which is already well known to the aircraft industry. The armature design reduces the effects of shock, vibration and acceleration; a spring type armature hinge eliminates backlash, friction and risk of displacement.



SERIES 595

COIL:
In all standard voltages up to 110 V, D.C. Working range between plus 10% and minus 20% of rating.

CONTACT COMBINATIONS: Up to 2 pole changeover.

BASE:
International octal.

WEIGHT: 3¼ ozs.

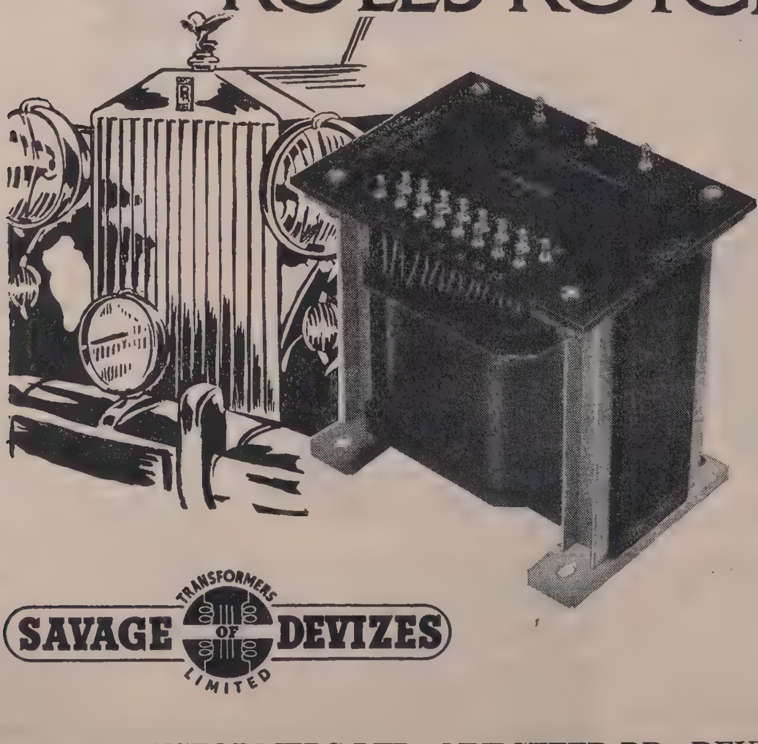
DIMENSIONS:
Diameter—1 $\frac{11}{16}$″ overall.
Height—2⅝″ overall.

SERIES 595H

Telephones Newmarket 3181-2-3 Telegrams: Magnetic, Newmarket

MAGNETIC DEVICES LTD
NEWMARKET

ITS THE ROLLS-ROYCE PERFECTION THAT COUNTS



MASSICORE SAVAGE

DEAR SIRS,

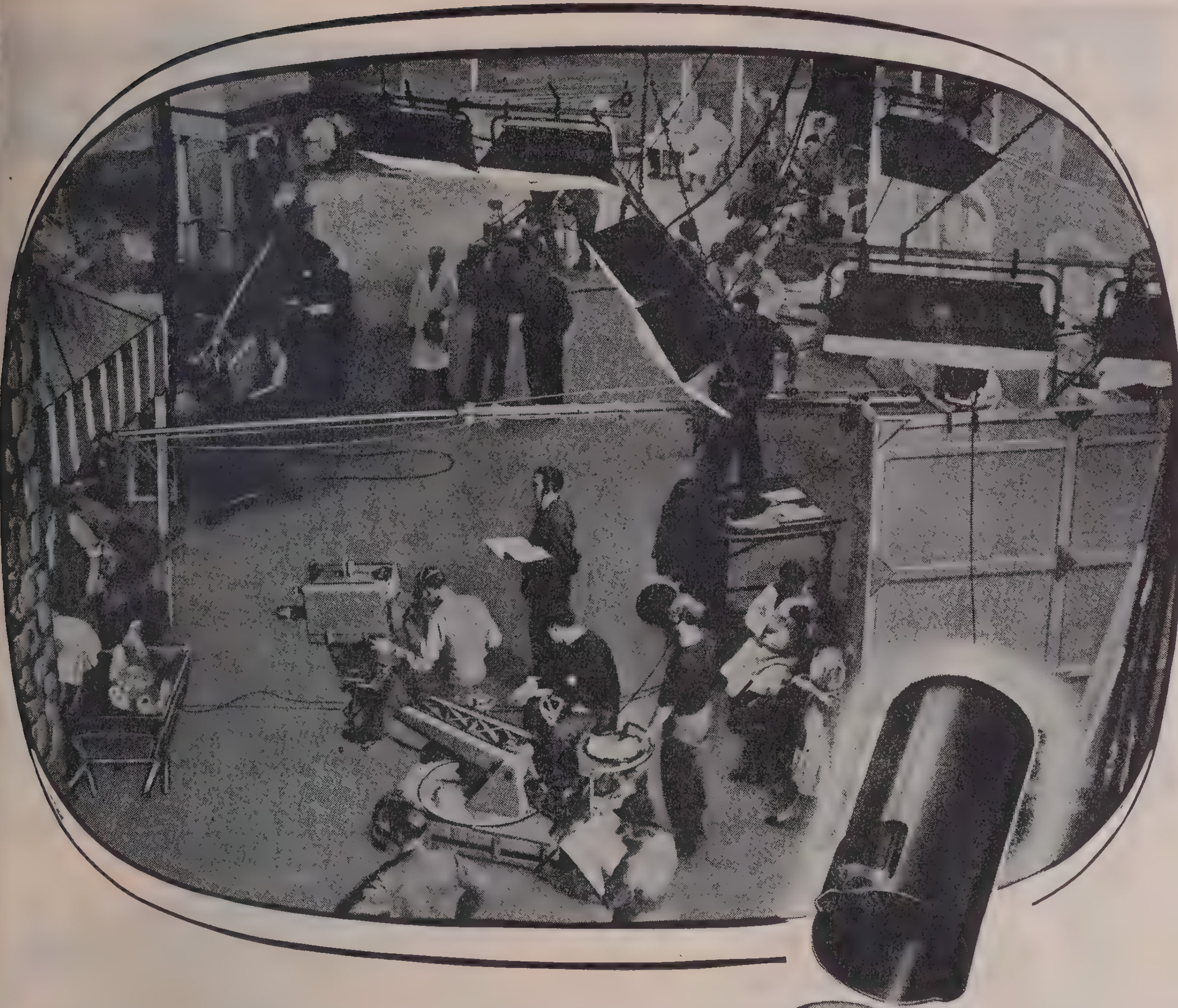
I am a customer, having two of your Transformers, which I believe are the best in the world. . . .

G.L.A.B., Washington, U.S.A.

Savage "Massicore" Transformers are meeting more exacting electronic needs than ever before—including those of Rolls-Royce, makers of the best *car* in the world.

SAVAGE OF DEVIZES TRANSFORMERS LIMITED

SAVAGE TRANSFORMERS LTD., NURSTEED RD., DEVIZES, WILTS. Telephone: DEVIZES 932



Why Ediswan Clix P.T.F.E. Valveholders are widely used in B.B.C. Television equipment

Large quantities of Ediswan Clix P.T.F.E. Valveholders are used in B.B.C. Television equipment.
Only the combination of the finest insulation—P.T.F.E., the most efficient contact material—Berylium copper—and Ediswan Clix design and manufacture can match the requirements of efficiency and reliability in this and all other stringent valveholder applications.
Ediswan Clix P.T.F.E. Valveholders are fully type approved for Services Grade 1, Class 1 conditions. Full details of these valveholders and other components in the Ediswan range are given in catalogue CR. 1681. Manufacturers and Development Groups may have a copy on request.

EDISWAN

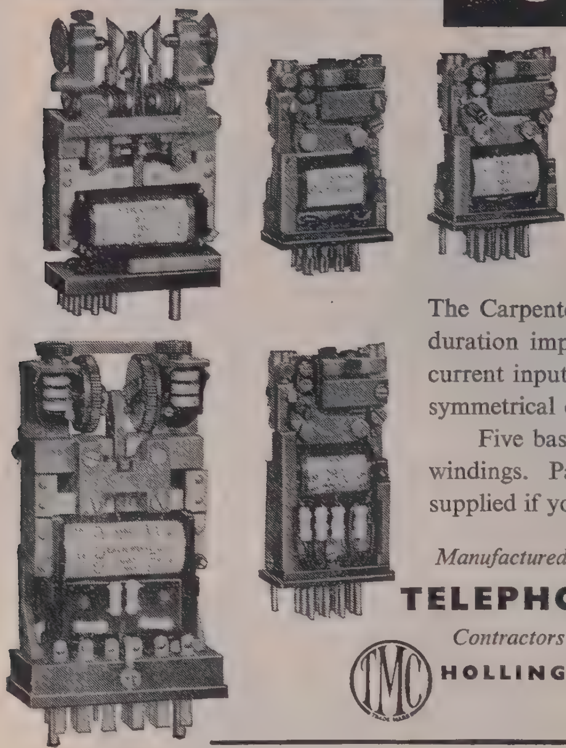
CLIX

RADIO, TELEVISION & ELECTRONIC COMPONENTS

THE EDISON SWAN ELECTRIC COMPANY LIMITED, *Member of the A.E.I. Group of Companies*
155 Charing Cross Road, London, W.C.2 and Branches. *Telephone: Gerrard* 8660. *Telegrams: Ediswan Westcent, London*

CR3

CARPENTER POLARIZED RELAYS



have these outstanding features

**HIGH OPERATIONAL SPEED • HIGH SENSITIVITY
FREEDOM FROM CONTACT REBOUND
NO POSITIONAL ERROR • HIGH CONTACT PRESSURES
EXCEPTIONAL THERMAL STABILITY • EASE OF
ADJUSTMENT • ACCURACY OF SIGNAL REPETITION**

The Carpenter Polarized Relay will respond to weak, ill-defined or short-duration impulses of differing polarity, or it will follow weak alternating current inputs of high frequencies and so provide a continuously operating symmetrical changeover switch between two different sources.

Five basic types are available with a wide range of single and multiple windings. Particulars of the type best suited to your purpose will be gladly supplied if you will send us your circuit details.

Manufactured by the sole licensees

TELEPHONE MANUFACTURING CO. LTD

Contractors to Governments of the British Commonwealth and other Nations

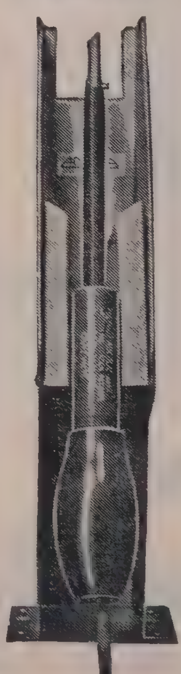
TMC

HOLLINGSWORTH WORKS, DULWICH, LONDON, SE21

Telephone GIPsy Hill 2211

ADCOLA
PRODUCTS LIMITED
(Regd. Trade Mark)

SOLDERING INSTRUMENTS & ALL ALLIED EQUIPMENT



ASSURES

SOUND JOINTS FOR SOUND EQUIPMENT

BIT SIZES
$\frac{1}{8}''$ to a $\frac{1}{4}''$

VOLT RANGES FROM
6/7 to 230/50 VOLTS

WITH NO EXTRA COST FOR LOW VOLTAGES

ADCOLA
PRODUCTS LTD.
Head Office & Sales
GAUDEN ROAD
CLAPHAM HIGH St.
LONDON, S.W.4

TELEPHONES
MACaulay 4272

Always in demand for precision work

Lewcos

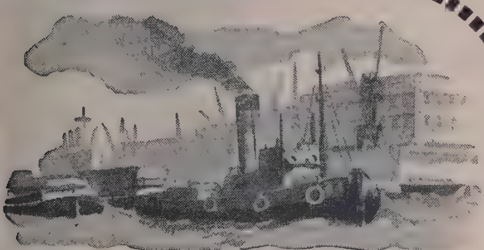
Insulated Resistance Wires with standard coverings of cotton, silk, glass, asbestos, standard enamel and synthetic enamel are supplied over a large range of sizes.

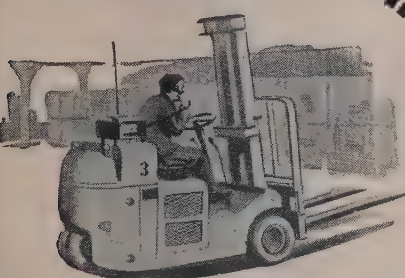
Send for leaflet LB11

THE LONDON ELECTRIC WIRE COMPANY
AND SMITHS, LIMITED
CHURCH RD., LEYTON, LONDON, E.10

Incorporating Frederick Smith & Company
Associated with The Liverpool Electric Cable Co. Ltd. and Vactite Wire Co. Ltd.

PYE REPORTER





LOW COST—SMALL SIZE
LOW BATTERY DRAIN
MOBILE RADIO TELEPHONE

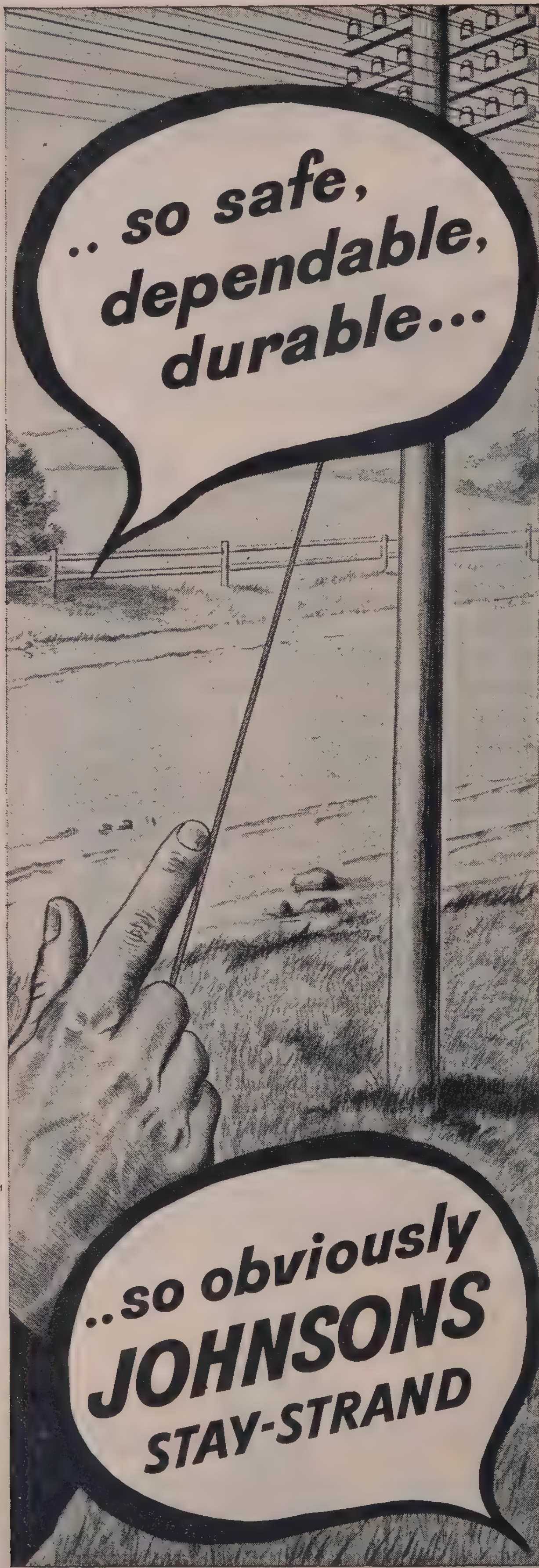
On airfields, in dockyards, in civil engineering and industry, operations reach their peak of speed and efficiency with the Pye "Reporter". All mobile personnel over a wide area can be contacted instantly through this compact and economical equipment which is quickly fitted to any vehicle. Control is simpler and swifter because wasted time and misunderstandings are eliminated. The Reporter enables four vehicles to do the work of five.

Pye Telecommunications are the largest suppliers of commercial 2-way radio in Europe and supply over 75% of British requirements.

PYE Telecommunications
CAMBRIDGE ENGLAND

Pye (New Zealand) Ltd. Auckland C.I., New Zealand	Pye Canada Ltd. Ajax, Canada	Pye-Electronic Pty., Ltd. Melbourne, Australia	Pye (Ireland), Ltd. Dublin, Eire
Pye Radio & Television (Pty.) Ltd. Johannesburg South Africa	Pye Limited Mexico City	Pye Limited Tucuman 829 Buenos Aires	Pye Corporation of America 270, Park Avenue New York

PYE LIMITED • • CAMBRIDGE • • ENGLAND
PHONE TEVERSHAM 311



RICHARD JOHNSON & NEPHEW LTD. FORGE LANE MANCHESTER 11

ZENITH

(REGD. TRADE-MARK)

Automatic

VOLTAGE REGULATORS

with Electronic Control

Essential for many purposes where a constant pre-determined supply voltage is required, these Regulators are designed to give a controlled output voltage within 1% with input voltage variations up to plus or minus 10%. Manufactured for single- and three-phase loads from 5 up to 23 kVA per phase.

Illustrated brochure free on request.

The ZENITH ELECTRIC CO. Ltd

ZENITH WORKS, VILLIERS ROAD, WILLESDEN GREEN LONDON, N.W.2

Telephone: WILlesden 6581–5 *Telegrams:* Voltaohm, Norphone, London

MANUFACTURERS OF ELECTRICAL ENGINEERING PRODUCTS INCLUDING RADIO AND TELEVISION COMPONENTS

AVO (REGD. TRADE MARK)

Precision ELECTRICAL TESTING INSTRUMENTS

The Model 7 Universal AVOMETER

A multi-range A.C./D.C. Electrical Measuring Instrument of B.S. 1st grade accuracy, providing fifty ranges of readings on a 5-inch hand-calibrated scale fitted with an anti-parallax mirror.

The meter will differentiate between A.C. and D.C. supply, the switching being electrically interlocked. The total resistance of the meter is 500,000 ohms.

CURRENT: A.C. and D.C. 0 to 10 amps.

VOLTAGE: A.C. and D.C. 0 to 1,000 volts

RESISTANCE: Up to 40 megohms.

CAPACITY: .01 to 20μF.

AUDIO-FREQUENCY POWER OUTPUT: 0–2 watts.

DECIBELS: —25Db. to + 16 Db.

The instrument is self-contained, compact and portable, simple to operate and almost impossible to damage electrically. It is protected by an automatic cut-out against damage through severe overload.

Power Factor and Power can be measured in A.C. circuits by means of an external accessory (the Universal AvoMeter Power Factor and Wattage Unit).

Other accessories are available for extending the wide ranges of measurements quoted above.

Write for fully descriptive pamphlet.

Size: 8 ins. × 7½ ins. × 4½ ins.
Weight: 6¾ lbs.

£19 : 10s.

Sole Proprietors and Manufacturers

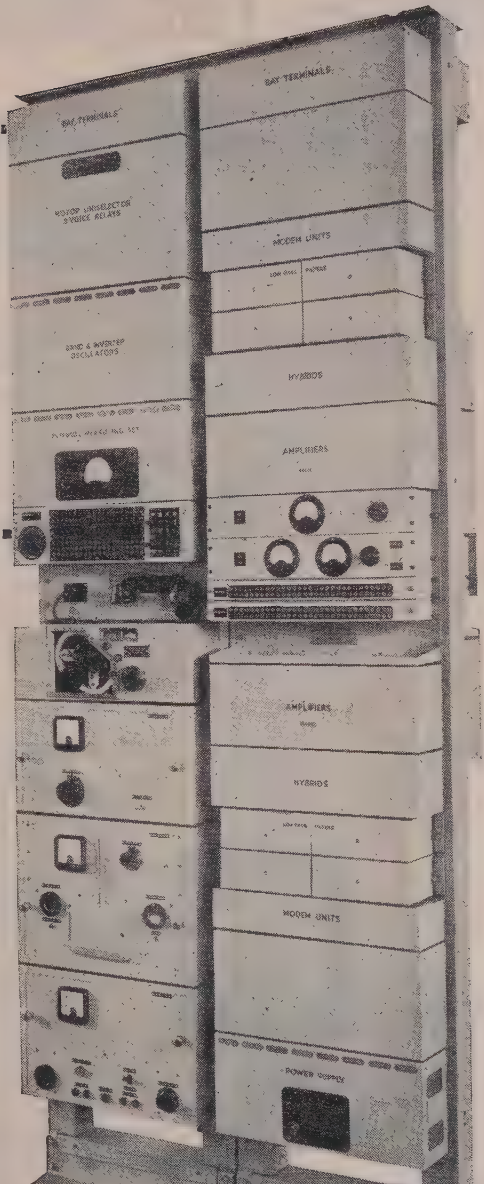
THE AUTOMATIC COIL WINDER & ELECTRICAL EQUIPMENT CO. LTD.

92–96 Vauxhall Bridge Road, London S.W.1 *Tel. VICTORIA 3404/9*

MARCONI–SIEMENS

Five Band Split Privacy Radio Telephone Equipment

(Type HW 12)



This equipment, which may be switched in or out of use at the radio terminal, provides a very high degree of privacy for speech on a radio-telephone circuit by:-
(1) splitting the speech band of 250-3000 c/s into five sub-bands of 550 c/s and recombining them in different relative positions,
(2) inverting the frequency range of any one or more of the sub-bands, and
(3) rearranging the combination of the sub-bands simultaneously at both ends of the radio-circuit in accordance with a pre-arranged sequence at controlled intervals between 4 and 20 seconds.
The resulting speech band, which modulates the transmitter, is unintelligible and the frequent regrouping of the sub-bands, with or without inversion precludes any simple method of interception. A reversal of the process at the distant terminal restores the original speech. The processes involved are reversible, thus common channel equipment can be used for both transmission and reception. Amplifiers in the privacy path compensate for the losses in band splitting and recombining. The simultaneous switching system, operates by means of relays under the control of a synchronous motor driven by a high precision crystal oscillator, this does away with the need for a transmitter pilot tone.

Marconi THE LINK BETWEEN RADIO AND LINE COMMUNICATIONS SIEMENS BROTHERS

Full technical details of this and other Marconi-Siemens equipment, which provides completely integrated radio and line telegraph and telephone systems may be obtained from either

MARCONI'S WIRELESS TELEGRAPH COMPANY LIMITED, CHELMSFORD, ESSEX
OR SIEMENS BROTHERS & CO., LIMITED, WOOLWICH, LONDON, S.E.18

MS2

AIRMEC

FREQUENCY STANDARD TYPE 761

THE AIRMEC FREQUENCY STANDARD TYPE 761 has been designed to fill the need for a self-contained frequency standard of moderate cost and high accuracy. It incorporates an oscilloscope for visual frequency comparison, and a beating circuit and loudspeaker for aural checking. A synchronous clock, driven from a voltage of standard frequency provides a time standard and enables long time stability checks to be made.

- **Master Oscillator:** Crystal-controlled at a frequency of 100 kc/s. The crystal is maintained at a constant temperature by an oven.
- **Outputs:** Outputs are provided at 100 c/s, 1 kc/s, 10 kc/s, 100 kc/s, and 1 Mc/s.
- **Waveform:** The above outputs are available, simultaneously with sinusoidal or pulse waveform from separate plugs.
- **Stability:** Four hours after switching on a short term stability of better than 1 part in 10^6 is obtained.

Full details of this or any other Airmec instrument will be forwarded gladly upon request.

AIRMEC LIMITED HIGH WYCOMBE BUCKINGHAMSHIRE ENGLAND

Telephone: High Wycombe 2060 *Cables:* Airmec, High Wycombe

PHILIPS

In Science and Industry alike . . .

among technicians, manufacturers and those engaged in the sale of electrical products — as well as among the public at large, the Philips emblem is accepted throughout the World as a symbol of quality and dependability.

PHILIPS ELECTRICAL LTD

CENTURY HOUSE, SHAFTESBURY AVENUE, LONDON, W.C.2

RADIO & TELEVISION RECEIVERS · RADIOGRAMS & RECORD PLAYERS · GRAMOPHONE RECORDS · TUNGSTEN, FLUORESCENT, BLENDED AND DISCHARGE LAMPS & LIGHTING EQUIPMENT · 'PHILISHAVE' ELECTRIC DRY SHAVERS · 'PHOTOFLUX' FLASHBULBS · HIGH FREQUENCY HEATING GENERATORS · X-RAY EQUIPMENT FOR ALL PURPOSES · ELECTRO-MEDICAL APPARATUS · HEAT THERAPY APPARATUS · ARC & RESISTANCE WELDING PLANT AND ELECTRODES · ELECTRONIC MEASURING INSTRUMENTS · MAGNETIC FILTERS · BATTERY CHARGERS AND RECTIFIERS · SOUND AMPLIFYING INSTALLATIONS · CINEMA PROJECTORS · TAPE RECORDERS

(P23)

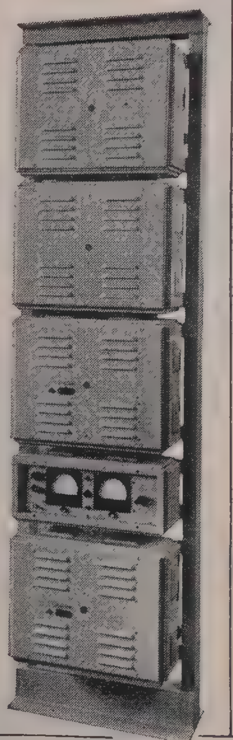
The Invisible link with the Isolated Community

V.H.F.

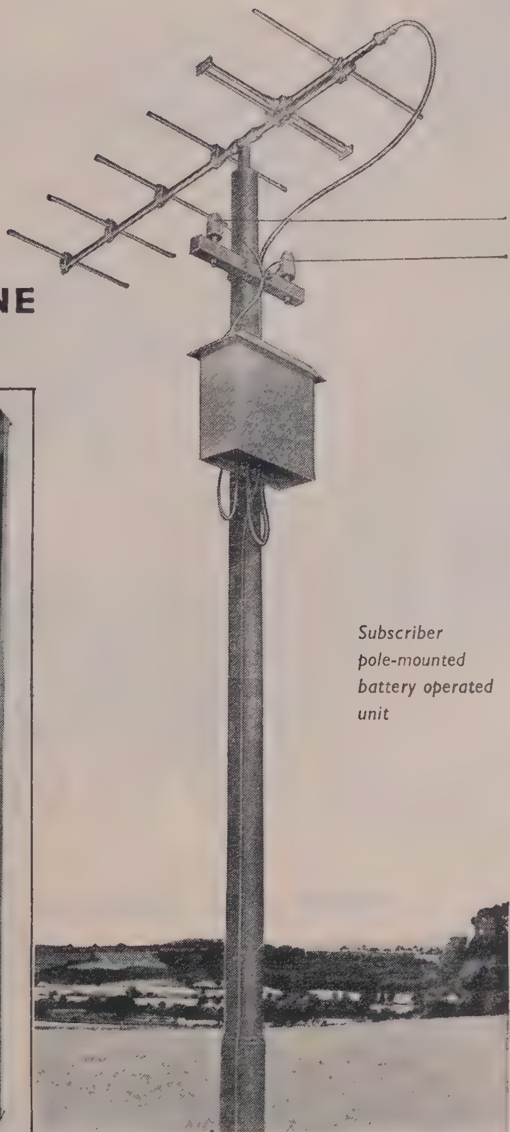
RADIO TELEPHONE

- ★ No Change in Normal Telephone Operating Procedure
- ★ Mains or Battery Operation
- ★ Signalling Units for All Types of Circuit

The V.H.F. link provides the most practical means of direct communication between isolated communities in all areas where the nature of the terrain or distance involved preclude the use of open wires for junction or subscribers' lines. Dialling facilities can be employed, and the radio equipment can be interposed in a standard line circuit in any part of a telephone system without modification to switching equipment.



Exchange equipment bay



Subscriber pole-mounted battery operated unit

AUTOMATIC TELEPHONE & ELECTRIC CO. LTD.

Radio & Transmission Division, Strowger House, Arundel Street, London, W.C.2
'Phone: TEMple Bar 9262. 'Grams: Strowgerex London.

AT 14511-Bx107

SIEMENS R.I.N.D. CABLE

A New Development

The R.I.N.D. Cable is RADIALLY IMPREGNATED under strictly controlled conditions thereby presenting a new technique in cable manufacture. The paper insulation is thoroughly dried and impregnated, but contains no superfluous compound. Dimensions in accordance with B.S.480: 1954 specification for normal belted type cables.

R.I.N.D. cable is suitable for general use, but is indispensable for situations where steep gradients and high ambient temperatures are encountered. No special jointing method is necessary, standard joint boxes are used.

SIEMENS BROTHERS

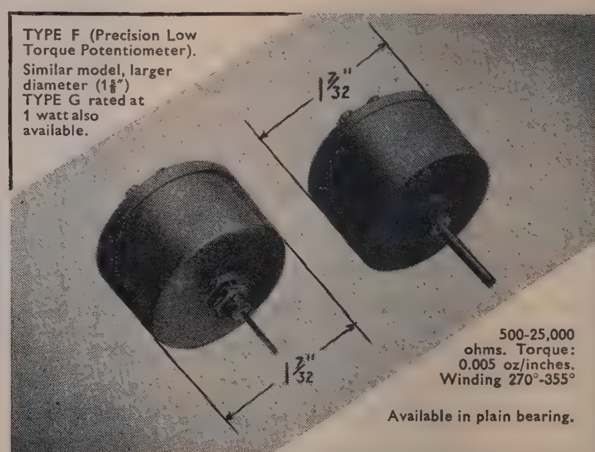
SIEMENS R.I.N.D. CABLE . . .

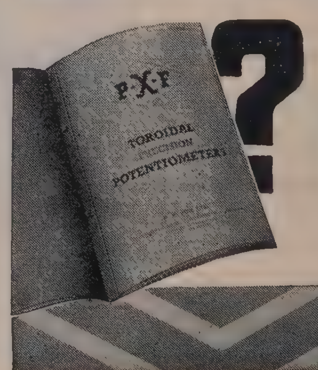
- *Passes Drainage Test in B.S.480/1954*
- *Has no critical temperature*
- *Gives no overload troubles*
- *Causes no difficulty with bending*
- *For voltages up to and including 11,000*

SIEMENS BROTHERS & CO. LIMITED,
WOOLWICH, LONDON, S.E.18.
Member of the A.E.I. Group of Companies

TOROIDAL POTENTIOMETERS

Ceramic Insulation only—and approved for Tropical conditions. Complete Ceramic Rings for strength. Also a large range of precision Toroidal-wound Potentiometers and Helical Potentiometers, 3 and 10 turn.





Have *you* a copy of this catalogue? If not, write for list No. 215

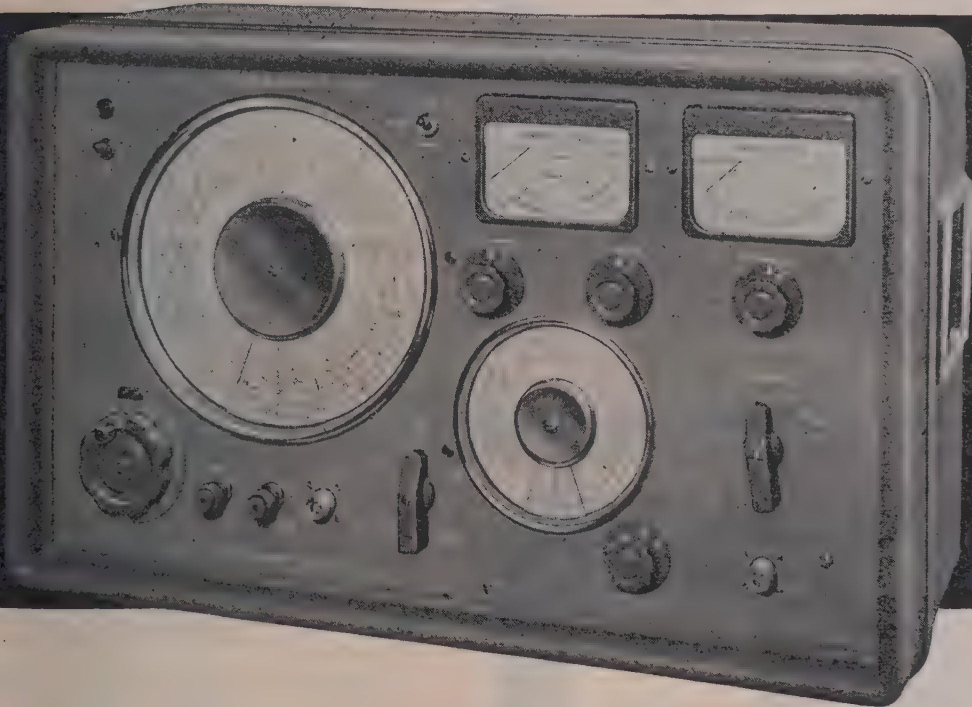
P · X · FOX LIMITED
HAWKSWORTH ROAD
HORSFORTH · YORKS

Tel: Horsforth 2831/2
Grams: Toroidal, Leeds

THE NEW

MARCONI SIGNAL GENERATOR

TYPE TF 801B



FREQUENCY RANGE 10 to 500 Mc/s

NORMAL OUTPUT: 0.1 μV to 0.5 volt, c.w. or modulated to 30%, available under all conditions ; 90% modulation at most carrier frequencies.
HIGH OUTPUT: up to 2 volts c.w., or 1 volt modulated, at most carrier frequencies. Sine a.m. : 1000-c/s internal or 30 c/s to 20 kc/s external. For external pulse modulation, r.f. bandwidth extends to 3 Mc/s. Output impedance : 50 ohms. A.C. mains operated.

MARCONI INSTRUMENTS

SIGNAL GENERATORS · BRIDGES · VALVE VOLTMETERS · Q METERS · WAVEMETERS
FREQUENCY STANDARDS · WAVE ANALYSERS · BEAT FREQUENCY OSCILLATORS

MARCONI INSTRUMENTS LTD · ST. ALBANS · HERTS · TELEPHONE: ST. ALBANS 6160/9

30 Albion St., Kingston-upon-Hull. Phone: Hull Central 16144. 19 The Parade, Leamington Spa. Phone: 1408

Managing Agents in Export:

MARCONI'S WIRELESS TELEGRAPH COMPANY LIMITED · MARCONI HOUSE · STRAND · LONDON · W.C.2

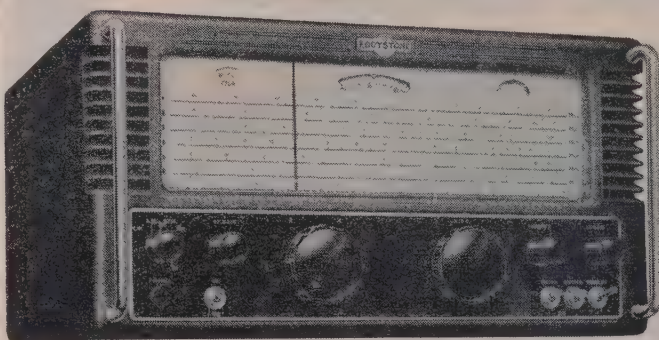
TC 68

EDDYSTONE

V.H.F. COMMUNICATIONS RECEIVERS

MODEL 770R – 19 Mc/s. to 165 Mc/s. CONTINUOUS
MODEL 770U – 150 Mc/s. to 500 Mc/s. CONTINUOUS

FOR MONITORING, FIELD TESTS, LABORATORY PURPOSES ETC.

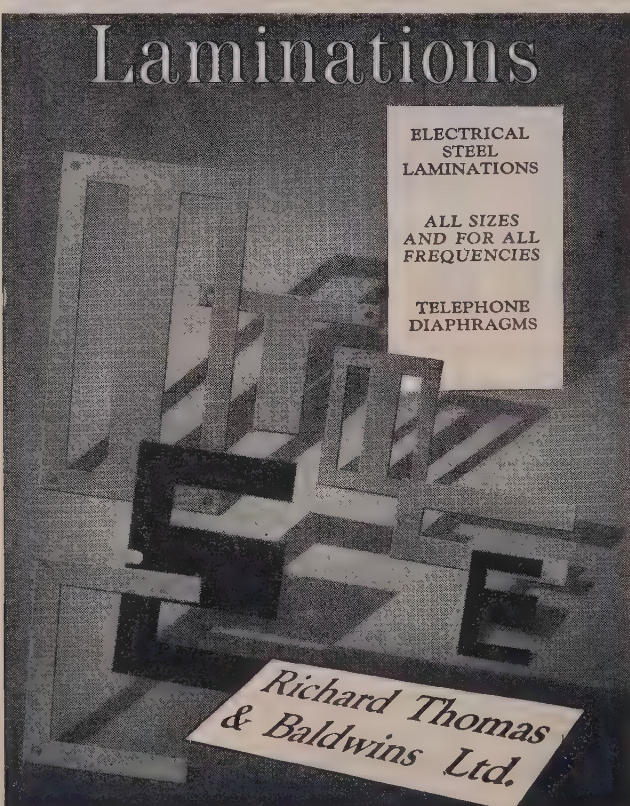


- Highly efficient signal frequency circuits.
- Substantial diecast rotary coil turret.
- Excellent frequency stability and selectivity.
- Accurate re-setting and ease of handling.
- High sensitivity and excellent signal-to-noise ratio.
- For AM and FM.
- Robust construction and outstanding reliability.
- "S" Meter. Noise Limiter.
- Preferred type valves.
- Finest workmanship.

- A. C. operation. 110-250 volts. 40-60 cycles.
- Dimensions 16¾″ x 15″ x 8¾″.
- Weight 60 lbs.

PLEASE WRITE FOR FULL SPECIFICATION TO THE MANUFACTURERS:

STRATTON & CO. LTD., ALVECHURCH ROAD, BIRMINGHAM, 31



A MEMBER OF THE RTSC GROUP
COOKLEY WORKS, BRIERLEY HILL, STAFFS.
Head Office: RTSC HOUSE, PARK ST, LONDON, W.1

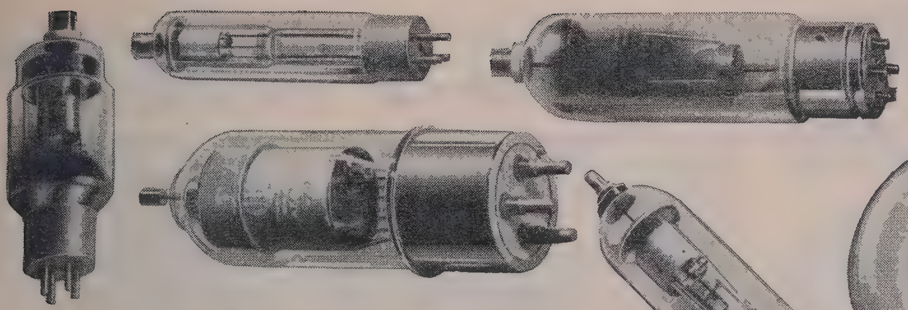
THE PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

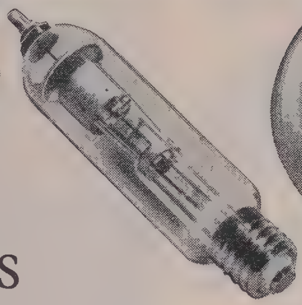
TEN-YEAR INDEX

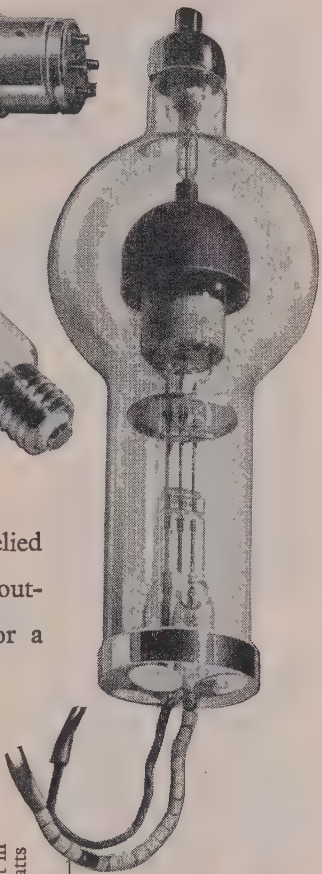
1942—1951

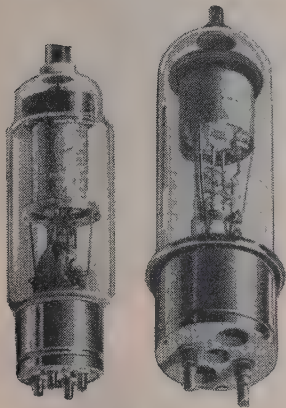
A TEN-YEAR INDEX to the *Journal of The Institution of Electrical Engineers* for the years 1942–48 and the *Proceedings* 1949–51 (vols. 89–98) can be obtained on application to the Secretary.

The published price is £1 5s. 0d. (post free), but any member of The Institution may have a copy at the reduced price of £1 (post free).



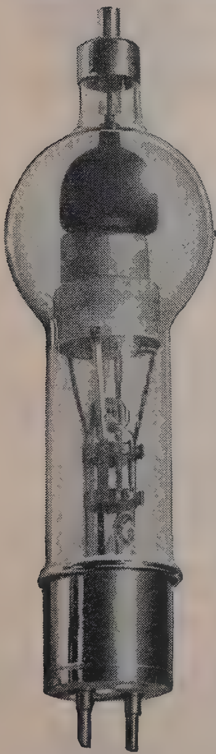




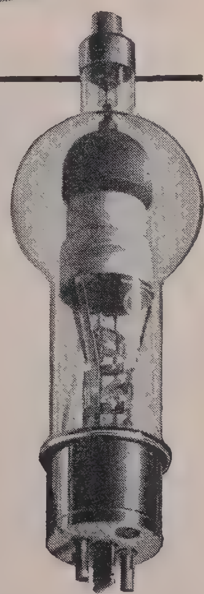


POWER RECTIFIERS

These Mercury Vapour and Xenon filled power rectifiers are robust and can be relied upon for long and satisfactory service. The output table shown herewith is calculated for a three-phase, full wave arrangement. Full technical data and operating instructions are available on request.



Tube Type	C.V. No.	Filling	Output in Kilowatts
AH 200	CV 1628	Mercury Vapour	90
AH 213	—	„ „	90
AH 201	—	„ „	7.5
AH 205	CV 2673	„ „	630
AH 211	CV 532	„ „	90
AH 221	CV 5	„ „	36
AH 217	—	„ „	36
AX 224	CV 1835	Xenon	7
AX 228	—	„	36
AX 230	CV 2518	„	36



'ENGLISH ELECTRIC'

ENGLISH ELECTRIC VALVE CO. LTD.

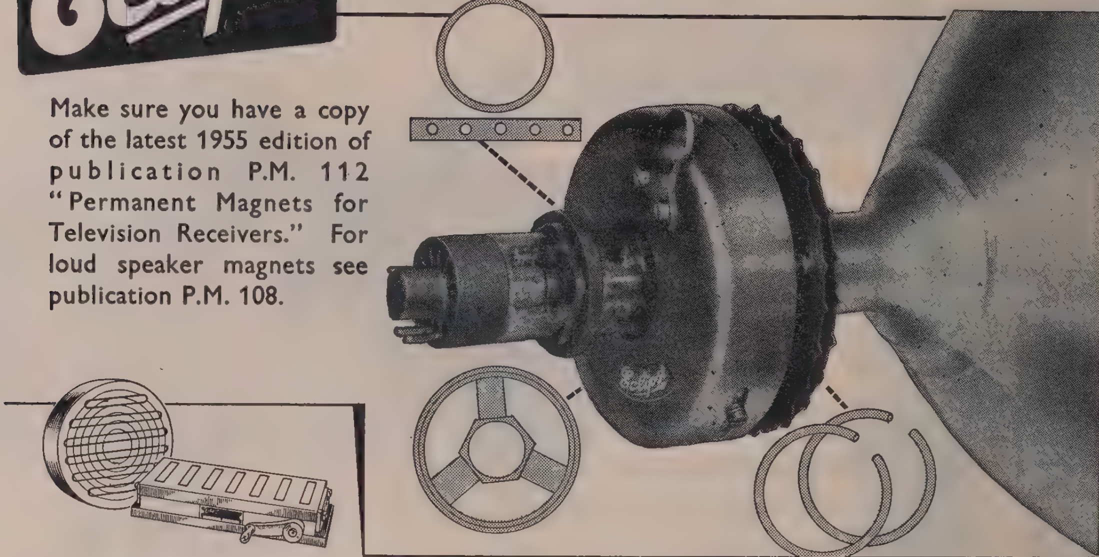
Waterhouse Lane, Chelmsford
Telephone: Chelmsford 3491

AP300/18

Eclipse

PERMANENT MAGNETS for TELEVISION RECEIVERS

Make sure you have a copy of the latest 1955 edition of publication P.M. 112 "Permanent Magnets for Television Receivers." For loud speaker magnets see publication P.M. 108.



MADE BY THE DESIGNERS & MANUFACTURERS OF ECLIPSE PERMANENT MAGNET CHUCKS

JAMES NEILL & CO. (SHEFFIELD) LTD.

SHEFFIELD 11 ENGLAND

M 4

THE PROFESSIONAL ENGINEERS APPOINTMENTS BUREAU

Incorporated under Limited Guarantee, 1947

(Licensed annually by the London County Council)

9 VICTORIA STREET, LONDON, S.W.1

The Principal Objectives of the Bureau are:

1. *To Aid Employers.*

To receive inquiries from employers seeking the services of qualified professional engineers and to advise such employers, if required, as to the qualifications desirable for the particular vacancy.

To submit to employers particulars of persons registered with the Bureau, the qualifications of whom seem to fit them for the appointments.

2. *To Aid Members of the Institutions of Civil, Mechanical and Electrical Engineers.*

To receive applications for registration for employment from engineers, who by reason of their engineering qualifications are Corporate or Non-Corporate Members of the Institution of Civil Engineers, The Institution of Mechanical Engineers, or The Institution of Electrical Engineers (excluding Associates of The Institution of Civil Engineers); and to charge and receive such registration fees and appointment fees as may from time to time be determined by the Bureau.

To give advice generally in matters relating to the employment of professional engineers.

EMPLOYERS

Employers of professional engineers who have vacancies to be filled, are invited to notify such vacancies to the Bureau and to furnish the following particulars:

(1) Title of post with description of duties involved.
(2) Age range. This should be made as wide as possible.
(3) Salary offered and supplementary remuneration, if any.
(4) Professional, technical and/or special qualifications, e.g. languages.
(5) Location of work.
(6) Could a successful candidate be assisted in finding suitable living accommodation locally?
(7) Does the post involve control of staff? If so, to what extent?
(8) Any other information which will assist in the selection of candidates.

Notifications of vacancies in Commonwealth countries are particularly welcome.

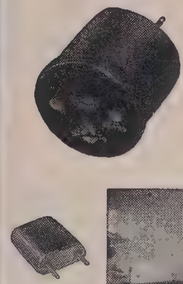
The Bureau deals with all classes of appointments ranging from practical training for Students to executive or administrative posts for senior Corporate Members.

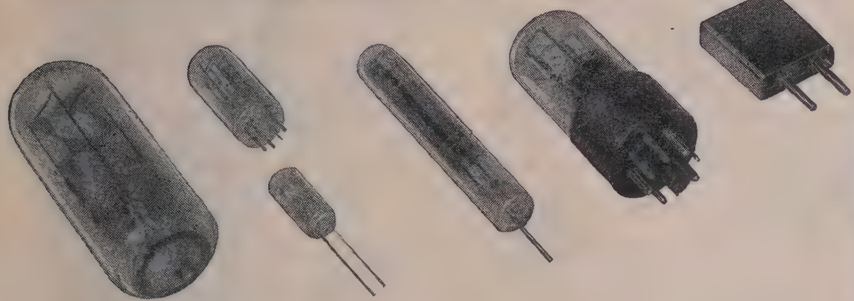
MEMBERS OF THE THREE INSTITUTIONS

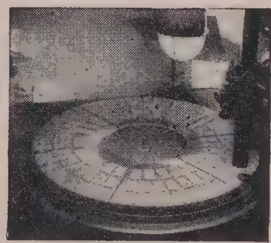
Members of the three Institutions who wish to register with the Bureau for employment may obtain the necessary forms on application to the Registrar. The grade of membership should be stated, and a stamped addressed foolscap envelope should be enclosed.

The Professional Engineers Appointments Bureau receives a number of applications for assistance from young graduate engineers, Civil, Mechanical and Electrical, who require practical experience as trainees. The Bureau would therefore be grateful to receive from firms advice of such vacancies as they occur, or when annual intakes are being considered.

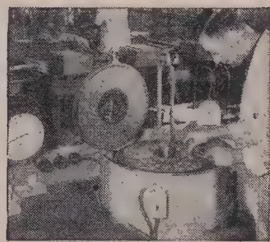
MARCONI CRYSTALS for Stability and Precision





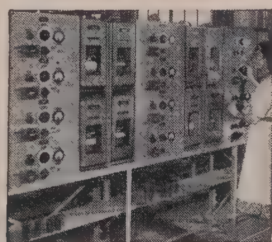


Surface Grinding

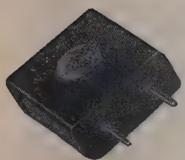


Planetary Lapping

Finishing to Frequency



Testing–Grid Current Recording



FREQUENCY RANGE 1·6 Kc/s TO 55 mc/s

The experience gained in manufacturing quartz crystals to the stringent requirements of our own apparatus and those of the Services, enables us to offer a comprehensive range of crystals covering the frequency band 1.6 Kc/s to 55 mc/s. Years of intensive research and development work in this field guarantee the reliability and quality of this Marconi Product.

Lifeline of communication

MARCONI

MARCONI'S WIRELESS TELEGRAPH CO. LTD., CHELMSFORD, ESSEX Telephone: Chelmsford 3221

CR 2

THERMIONIC VALVES

The First Fifty Years

A special publication of

THE INSTITUTION OF ELECTRICAL ENGINEERS

reporting the celebration of the Jubilee

of the thermionic valve

Price 9s.

Profusely illustrated, printed entirely on art paper and bound in substantial semi-stiff covers, the Institution publication entitled "Thermionic Valves—the First Fifty Years" made its appearance towards the end of June. As already announced, this book contains the lectures delivered by Sir Edward Appleton, Professor G. W. O. Howe and Dr. J. Thomson at the Jubilee Meeting in November, 1954, together with brief descriptions of about 250 valves—from Fleming's diodes to travelling-wave tubes—which were gathered together for the occasion.

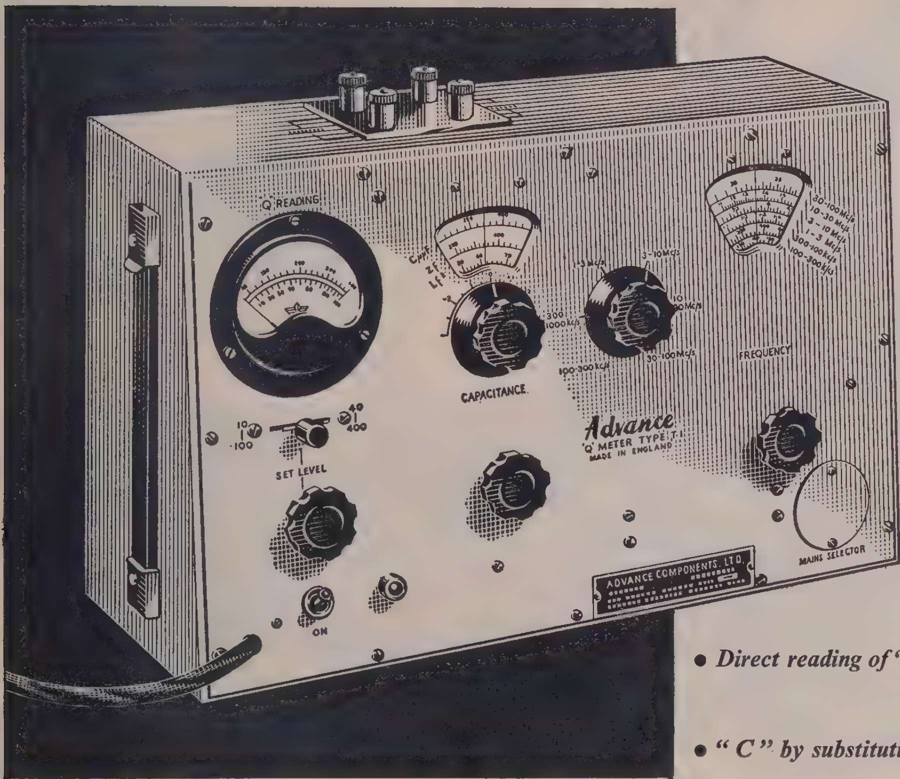
The size of the book is 76 pages, demy quarto, and its price (post free) is 9s. to the public, and 4s. to members of The Institution. The edition is limited and orders should be sent to the Secretary without delay.

Obtainable on application to the Secretary

THE INSTITUTION OF ELECTRICAL ENGINEERS

SAVOY PLACE, LONDON, W.C.2

ANNOUNCING THE Advance 'Q' METER



- *Direct reading of "Q" Range 10-400*
- *"C" by substitution.*
- *Rapid calculation of "L" and "Z"*
- *No "Set-Zero" problems*
- *Small and portable*

The ADVANCE "Q" Meter is different! It is small, portable and has an excellent specification—a useful addition to any electronic laboratory and well suited for production testing. Furthermore, it is offered at a price to suit all applications. With the T1, RF measurements can be made of "Q" inductance, impedance, capacitance and power factor at frequencies between 100 kc/s. and 100 Mc/s.

Full details in leaflet A/31 which we will be pleased to forward on your request.

TYPE T1 £55 NETT PRICE IN U.K.

ADVANCE COMPONENTS LTD. Marlowe Rd., Walthamstow, London, E.17 Tel: LARkswood 4366/7/8

Lines of communication . . .

In June 1920 the first advertised sound radio programme was broadcast from the Marconi transmitter at Chelmsford. To-day Marconi high or medium power transmitters and high power aerials are installed in every one of the B.B.C's television transmitter stations and Marconi television systems are being supplied to countries in North and South America, Europe and Asia.

TELEVISION CAMERAS
STUDIO EQUIPMENT
RADIO LINKS AND TRANSMITTERS
AERIAL SYSTEMS
COLOUR TELEVISION
INDUSTRIAL TELEVISION

MARCONI

Complete Television and Sound Broadcasting Systems

MARCONI'S WIRELESS TELEGRAPH COMPANY LIMITED · CHELMSFORD · ESSEX

LG 3

'Araldite'

epoxy casting resins

epoxy

epoxies

ite'

'Araldite'

epoxy resin adhesives

epoxy surface coating resins

'Araldite'

'Ara

ng resins

These versatile resins have a remarkable range of characteristics and uses.

They are used
- for bonding metals and ceramics.
- for potting and sealing electrical components.
- for producing glass cloth laminates.
- for producing jigs, fixtures, patterns and tools.
- as fillers for sheet metal work.
- as protective coatings for metal surfaces.

epoxy casting resins

'Araldite'

Araldite is a Registered Trade Mark

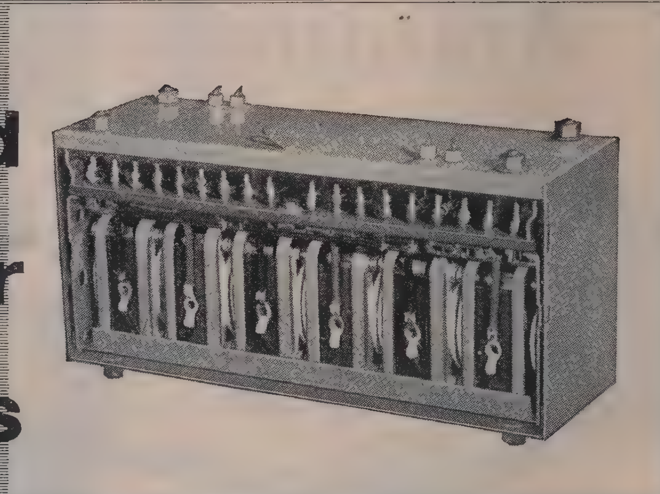
FULL DETAILS WILL BE SENT GLADLY ON REQUEST

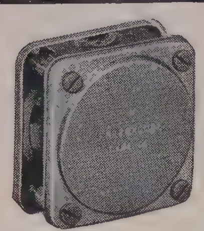
Aero Research Limited

A Ciba Company

Duxford Cambridge. Telephone : Sawston 187

AP 264-141A





Improved filter units

with Ferroxcube pot cores

1 *High performance combined with small size and light weight.*

2 *Designed and built to customers' individual requirements.*

3 *Long term stability, even under conditions of temperature variation.*

High quality electrical filter units built around Ferroxcube cores can now be supplied to communications equipment designers' individual specifications. These filter units have significant advantages over comparable types designed without the use of Ferroxcube, particularly in the frequency range 300 c/s to 500 kc/s. For audio frequencies the use of Ferroxcube cores permits the winding of compact coils with very high inductances. This results in a considerable reduction in the size and cost of the associated condensers and hence of the filter unit as a whole. The high Q values obtained for a given volume, especially above 10 kc/s, enable sharp cut off characteristics and low pass-band losses to be achieved, while negligible stray flux facilitates the production of compact and mechanically robust filters. Electrical filter units are among a number of high quality components now being made available by Mullard. Full details of the complete series of components will be gladly supplied upon request.

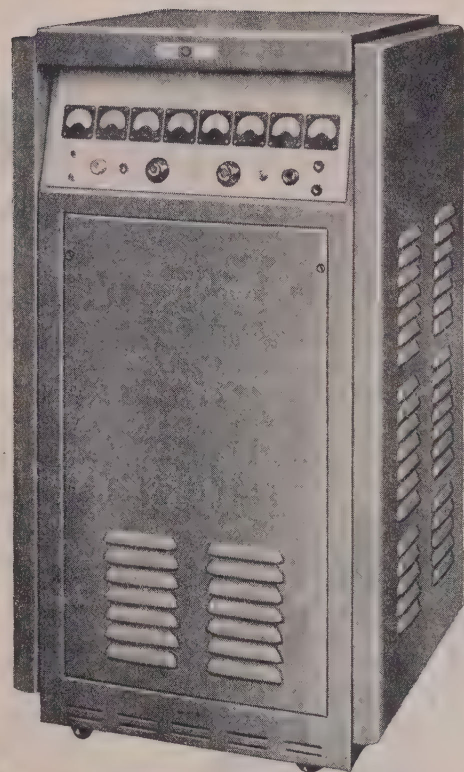
Mullard

'Ticonal' alloy permanent magnets
Magnadur permanent ceramic magnets
Ferroxcube ferro-magnetic cores.

MULLARD LTD · COMPONENTS DIVISION · CENTURY HOUSE · SHAFTESBURY AVENUE · WC2

(MI463A)

20 Mc/s FREQUENCY MONITOR



The Automatic Frequency Monitor (20 Mc/s) is but one of a series of high grade monitors now in course of manufacture for the accurate measurement of frequency.

Employing hard valve techniques throughout, it will measure any frequency in the range 10 c/s to 20 Mc/s to an accuracy within $\pm$ 1 part in 10^6.

The result, in decimal notation, is presented on eight panel mounted meters each scaled from 0 to 9 and the unknown frequency is automatically remeasured every few seconds.

This new equipment presents a considerable advance in frequency measuring techniques and apart from normal laboratory applications, is ideally suited for incorporation in production testing routines.

Full technical information on this and other frequency measuring equipment is available on request.

CINEMA TELEVISION LTD

A COMPANY WITHIN THE J. ARTHUR RANK ORGANISATION

WORSLEY BRIDGE ROAD · LONDON · S.E.26

HITHER GREEN 4600

SALES AND SERVICING AGENTS:

Hawnt & Co. Ltd., 59 Moor St. Birmingham, 4

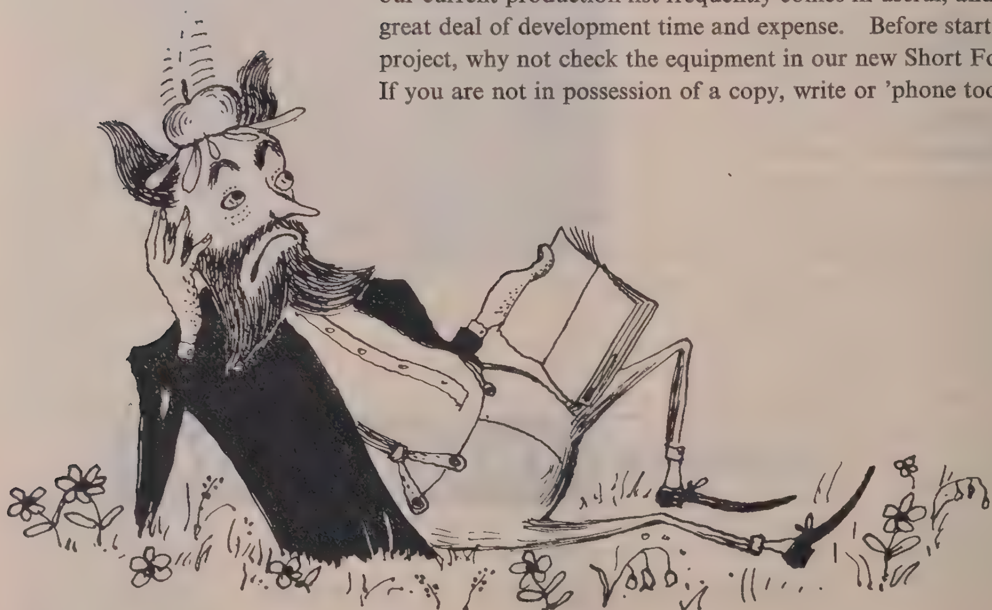
Atkins, Robertson & Whiteford Ltd., 100 Torrisdale Street. Glasgow, S. 2

F. C. Robinson & Partners Ltd., 122 Seymour Grove, Old Trafford, Manchester, 16

An apple a day . . .

keeps the Doctor awake; in fact, a violent blow on the head is just the stimulus required to knock his latest daydreams into shape. Catching the cat neatly in the midriff with his trusty philosopher's stone, Frittertime lights out on the trail of his panacea for all scientific ills.

We wouldn't dream of insinuating that any of your people rely on this kind of inspiration, but we know that a knowledge of the instruments on our current production list frequently comes in useful, and can save you a great deal of development time and expense. Before starting on that new project, why not check the equipment in our new Short Form Catalogue? If you are not in possession of a copy, write or 'phone today.



THE SOLARTRON ELECTRONIC GROUP LTD

RELIABILITY—UNDER OUR 12 MONTHS' GUARANTEE, COSTS HAVE NEVER EXCEEDED 0·2% OF SALES

Thames Ditton, Surrey
Telephone: EMBerbrook 5522
Cables: Solartron, Thames Ditton

The Institution is not, as a body, responsible for the opinions expressed by individual authors or speakers.
An example of the preferred form of bibliographical references will be found beneath the list of contents.

THE PROCEEDINGS OF
THE INSTITUTION OF ELECTRICAL ENGINEERS

EDITED UNDER THE SUPERINTENDENCE OF W. K. BRASHER, C.B.E., M.A., M.I.E.E., SECRETARY

VOL. 102. PART B. No. 5. SEPTEMBER 1955

1–52:621.3.018.3

Paper No. 1693 M
July 1954

THE MECHANISM OF SUB-HARMONIC GENERATION IN A FEEDBACK SYSTEM

By JOHN C. WEST, B.Sc., Ph.D., Associate Member, and JOHN L. DOUCE, M.Sc., Ph.D., Graduate.

(The paper was first received 25th February, and in revised form 30th April, 1954. It was published in July, 1954, and was read before the MEASUREMENTS SECTION 15th March, 1955.)

SUMMARY

A theory is presented for the mechanism of sub-harmonic generation in a non-linear feedback system. This phenomenon is the occurrence of components, in the output of a system, whose frequency is a fraction of the input frequency. This sub-harmonic component is regarded as an oscillation of the closed-loop feedback system in the Nyquist sense. The parameters of the system are modified to allow continuous oscillation by the action of the input signal on the non-linear element.

A connection is shown to exist between the open-loop frequency response and the conditions necessary for sub-harmonic oscillation. From this, sufficient limitations are obtained to ensure the absence of this phenomenon.

Experimental work on a system with a cubic characteristic is included.

LIST OF PRINCIPAL SYMBOLS

v_i = Instantaneous amplitude of the signal applied to the non-linear element.
v_o = Instantaneous amplitude of the output of the non-linear element corresponding to v_i.
ω = Angular frequency.
ϕ, δ = Phase angles.
a = Amplitude of the component of v_i of sub-harmonic frequency.
b = Amplitude of the component of v_i of fundamental frequency.
$G(j\omega)$ = Open-loop gain of the linear system.
$N(v_i)$ = Gain of the non-linear element.

(1) INTRODUCTION

Many types of non-linear feedback systems exist which, when subjected to certain sinusoidal inputs, produce an output whose lowest-frequency component is a fraction of the input frequency. A common example is the relaxation oscillator, which can be set in continuous oscillation at a frequency much lower than, but synchronized with, an input or control signal. An LCR resonant circuit, with an iron-cored inductor as the non-linear element, can be set in oscillation at a frequency near that of natural resonance by the application of a suitable alternating voltage of twice this frequency.[1] A moving-coil loudspeaker with non-uniform magnetic field[2] is another example of a device by which sub-multiple harmonics can be generated.

This phenomenon can also occur in a class of remote-position-control servo mechanisms where saturation of the amplifier or motor provides a suitable non-linear characteristic. In all these cases the lowest frequency present at the output is a sub-multiple of the input frequency, and is usually at or near the natural frequency of oscillation. For this reason the phenomenon is known as "sub-harmonic resonance."

Because of the resonant nature of the phenomenon, the sub-harmonic component is often of large amplitude, so that the output is considerably different from that obtained when the phenomenon does not occur. For systems in which the output is intended to reproduce the input as closely as possible, as in the motion of a loudspeaker cone or of a servo motor, or for systems intended to respond only to a band of frequencies, this phenomenon can introduce an intolerable distortion. Hence it is of importance to lay down conditions on the parameters of the system to ensure the absence of the phenomenon. It does not occur in linear systems, and must therefore depend on the non-linear characteristic involved in the system.

In a practical system, it may be inconvenient or impossible to modify the form of the non-linear element. In this case it is important to decide how to modify the linear elements to eliminate sub-harmonic resonance.

Conversely there may be fields of application for devices where the phenomenon is desirable. It will then be necessary to derive the form of a suitable non-linear element to ensure sub-harmonic generation.

Previous work has accounted for sub-harmonic resonance by the solution of a non-linear differential equation, in which the non-linear characteristic is expressed as a power series.[3] There are two objections to this method. Since the output of the non-linear element is expressed as a power series of the input, successive approximations to the correct solution converge rapidly only if the amount of non-linearity is small. Thus the method is inapplicable to many practical systems, where sharp changes of slope occur—e.g. saturation effects.

Drs. West and Douce are in the Servo-Mechanisms Laboratory, Electrical Engineering Department, University of Manchester.

The second and perhaps more important disadvantage is that the solution of a differential equation gives no physical picture of the mechanism of the phenomenon, or of any useful design criterion.

(2) THE MECHANISM OF SUB-HARMONIC GENERATION

The basic system considered is shown in Fig. 1, and contains a linear frequency-dependent portion of gain $\psi(j\omega)$ and a non-linear element whose gain $N(v_i)$ depends solely on the magnitude

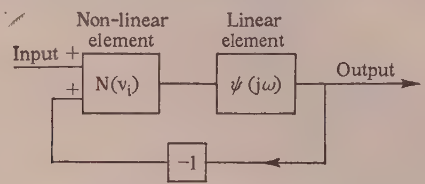


Fig. 1.—Block schematic of the general system.

of the input v_i to this element. When the non-linear element is replaced by an element of unity gain, the resultant linear system is assumed to be stable.

For the linear system, the superposition theorem is applicable, and hence the response of the closed-loop system to a given signal is independent of any other signal which is applied simultaneously. This is not the case for the system when a non-linear element is included in the feedback loop. In general, the response of the non-linear element to a given signal is modified if a further signal is applied at the same time. This further signal may be considered as varying the gain of the non-linear element for the primary input. Thus if the primary input $V_i\varepsilon^{j\omega t}$ is sinusoidal, the effect of an additional input is to vary the amplitude of the output $V_o\varepsilon^{j\omega t}$. A case of interest arises if the additional input is also sinusoidal, its frequency being an integral multiple of the primary input frequency. In this particular case, the output component of angular frequency ω may be modified in both amplitude and phase by the presence of the additional input. This phase change is due to the intermodulation products generated by the non-linear element.

Since the non-linear element is inserted in a feedback loop, it is evident that varying the gain of this element will affect the stability of the closed-loop system. For a range of sinusoidal inputs, the response of the non-linear element may be so modified that the closed loop is rendered oscillatory. Where the system is such that the open-loop frequency response, plotted on the complex plane, never cuts the negative real axis, the non-linear element must introduce a phase change for stable oscillations to occur. This implies that the frequency of oscillation must be simply related to the frequency of the primary input. The case of interest is where the frequency of oscillation is a sub-multiple of the input frequency, giving the phenomenon of sub-harmonic generation.

Thus sub-harmonic resonance is considered as an oscillation of the closed feedback loop. The loop gain necessary for this oscillation to be continuous is maintained by the action of the input on the non-linear element.

To discuss this theory in a quantitative manner, it is necessary to determine the response of the non-linear element to a sinusoidal signal with a further input of a harmonic frequency applied simultaneously. It is also necessary to consider the conditions required for the continuous oscillation of a non-linear feedback system.

The second point is evaluated using the "describing function" technique, introduced by Kochenburger[4] and Johnson.[9] A summary of this theory is given in the following Section.

(2.1) The Describing Function

A linear feedback system is incapable of maintaining co- tinuous oscillations if the open-loop frequency-response loc plotted on the complex plane, does not enclose the (−1, point.[10] If this response locus does enclose the −1 point, t system will oscillate with increasing amplitude. Only for t case where the locus passes through the −1 point will t oscillations be of constant amplitude.[5]

In a continuously oscillatory system involving a non-line element, the amplitude of the oscillations increases to son constant value. This final state of oscillation implies that tl overall open-loop gain is such that the locus now passes throug the −1 point. The response locus is now both amplitude- ar frequency-dependent, and the amplitude and frequency of th stable oscillations are the values necessary to give an open-loo gain of −1.

For the linear system, of open-loop $\psi(j\omega)$, the closed-loo gain is

$$G = \frac{\psi}{1+\psi} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad ($$

so that continuous constant-amplitude oscillations occur fo $\psi = -1$.

With a non-linear element in the loop of gain $N(v_i)$, whe $N(v_i)$ is the ratio of the amplitude of the fundamental of th output to the amplitude of the input sinusoid, the open-loo gain is $\psi(j\omega)N(v_i)$ and the closed-loop gain becomes

$$G(j\omega, v_i) = \frac{\psi(j\omega)N(v_i)}{1+\psi(j\omega)N(v_i)} \quad . \quad . \quad . \quad . \quad ($$

$$= \frac{\psi(j\omega)}{\frac{1}{N(v_i)}+\psi(j\omega)} \quad . \quad . \quad . \quad . \quad ($$

Thus oscillations of constant amplitude occur for

$$\psi(j\omega) = \frac{-1}{N(v_i)} \quad . \quad . \quad . \quad . \quad . \quad . \quad ($$

$-1/N(v_i)$ is termed the describing function of non-linearity.

To determine the amplitude and frequency of any stable osci lation of the closed loop, the frequency-dependent response the linear system, $\psi(j\omega)$, and the amplitude-dependent describir function, $-1/N(v_i)$, are plotted on the complex plane. Th point of intersection of the two loci gives the required informatio

For the discussion of sub-harmonic resonance, the describir function relates the sub-harmonic components before and aft the non-linear element. This describing function depends o the amplitude of the fundamental and sub-harmonic component on the phase relationship between them, and on the order the sub-harmonic considered.

It is shown in Section 8 that the sub-harmonic compone undergoes a phase shift in passing through the non-lineari only if its frequency is a simple fraction of the input frequenc Considering one particular sub-harmonic frequency, th describing function depends on three variables—the amplitud of the two components at the input to the non-linear eleme and the phase relationship between them. This input sign may be written

$$v_i = a\cos\omega t + b\cos(n\omega t + \phi) \quad . \quad . \quad . \quad ($$

As ϕ is varied the describing function, plotted on the compl plane, follows a closed curve symmetrical about the real axi Zero phase change occurs for $\phi = 0$ and $\phi = \pi$.

For the class of systems considered, b depends almost entire

on the amplitude of the primary input to the system, i.e. there is elatively little amplitude of this frequency component in the eedback signal in the frequency range considered. Thus b may e considered independent of the amplitude of the sub-harmonic omponent.

Thus for a given amplitude of input to the system, the lescribing function is a function of the two variables a and ϕ. A particular hard-spring non-linear element has a describing unction as shown in Fig. 2.

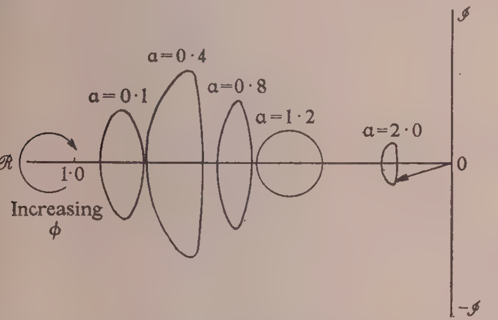


Fig. 2.—Describing-function loci of the cubic non-linearity $v_0 = v_i^3$ for variable ϕ.
$b = 0{\cdot}3$.

(3) A SPECIFIC NON-LINEARITY

As an example of the method of procedure, a non-linear lement whose output is proportional to the cube of the input considered,

e.
$$v_0 = v_i^3 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

his example is taken because of the ease with which a complete athematical treatment can be given, and it in no way provides limitation on the range of application of the theory.

Let the input be of the form

$$v_i = a \cos \omega t + b \cos (n\omega t + \phi) \quad . \quad . \quad . \quad (7)$$

here n is greater than unity but is not necessarily an integer,

$$= a^3 \cos^3 \omega t + 3a^2b \cos^2 \omega t \cos (n\omega t + \phi)$$
$$+ 3ab^2 \cos \omega t \cos^2 (n\omega t + \phi) + b^3 \cos^3 (n\omega t + \phi)$$
$$= \left(\frac{3a^3}{4} + \frac{3ab^2}{2}\right) \cos \omega t + \frac{a^3}{4} \cos 3\omega t$$
$$+ \left(\frac{3a^2b}{2} + \frac{3b^3}{4}\right) \cos (n\omega t + \phi) + \frac{3a^2b}{4}\{\cos [(n+2)\omega t + \phi]$$
$$+ \cos [(n-2)\omega t + \phi]\} + \frac{3ab^2}{4}\{\cos [(2n+1)\omega t + \phi]$$
$$+ \cos [(2n-1)\omega t + \phi]\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

qn. (8) may be rewritten with successive terms increasing with equency as

$$= \left(\frac{3a^3}{4} + \frac{3ab^2}{2}\right) \cos \omega t + \frac{3a^2b}{4} \cos [(n-2)\omega t + \phi]$$
$$+ \frac{3ab^2}{4} \cos [(2n-1)\omega t + \phi] + \frac{a^3}{4} \cos 3\omega t$$
$$\text{and terms in } n\omega t \text{ and higher} \quad . \quad (9)$$

It follows from eqn. (9) that there is a phase change in the low-frequency component of the output only if the second or third term involves $\sin \omega t$. For this it is necessary for

$$n - 2 = 1 \text{ or } 2n - 1 = 1 \quad . \quad . \quad . \quad . \quad (10)$$

i.e. $n = 3$, or $n = 1$ which is neglected.

Substituting $n = 3$ gives

$$v_o = \left(\frac{3a^3}{4} + \frac{3ab^2}{2}\right) \cos \omega t + \frac{3a^2b}{4} \cos (\omega t + \phi)$$
$$\text{and terms in } 3\omega t, \text{ etc.} \quad . \quad (11)$$

The information required is the amplitude and phase change of the low-frequency component in the output, the amplitude of the fundamental, of frequency $3\omega/2\pi$, being regarded as predetermined.

Let the phase change be δ. From eqn. (11) it follows that

$$\tan \delta = \frac{-ab \sin \phi}{a^2 + 2b^2 + ab \cos \phi} \quad . \quad . \quad . \quad (12)$$

As ϕ is varied, it may be shown that

$$(\tan \delta)_{max} = \frac{\pm ab}{\sqrt{(a^4 + 4b^4 + 3a^2b^2)}} \quad . \quad . \quad . \quad (13)$$

If b is maintained constant and a is varied, the maximum value of $\tan \delta$ occurs for $a = \sqrt{(2)}b$, for which $\delta = \pm 21°$.

Thus the describing function for this non-linearity always lies

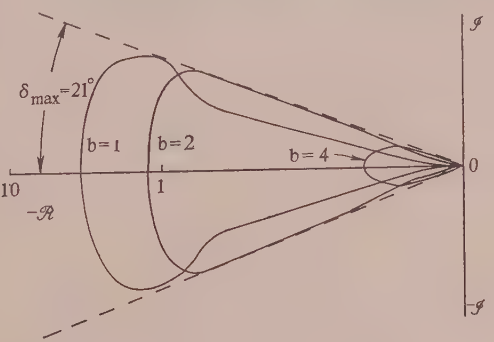


Fig. 3.—Boundaries of the describing function.

in the area bounded by the lines $\delta = \pm 21°$, as shown in Fig. 3. This Figure shows the envelope of the describing function for several values of b.

(3.1) The Experimental Non-Linear Element

Since it is relatively simple to derive the theoretical response of the non-linearity previously considered, an experimental unit having an approximately cubic response has been constructed.

The experimental non-linearity is basically an anode-follower, with resistive ratio arms.[6] Diode switches are added, so that as the input amplitude increases the input resistance decreases, and hence the gain of the element increases.

A total of 24 diodes are used (see Fig. 4), and the input resistances are calculated so that each diode changes the gain of the unit by a factor of $1{\cdot}5$. This is found to be a satisfactory compromise between the opposing claims of a smooth characteristic and a large total variation of gain.

The non-linear element is to be inserted in a d.c. feedback loop, and to ensure that the input is of the form assumed eqn. (7), i.e. with no steady applied voltage, the input is applied through an RC network of large time-constant. This introduces negligible phase shift over the working range of frequencies. A d.c. path is provided by the high resistance R_0 (see Fig. 4) in

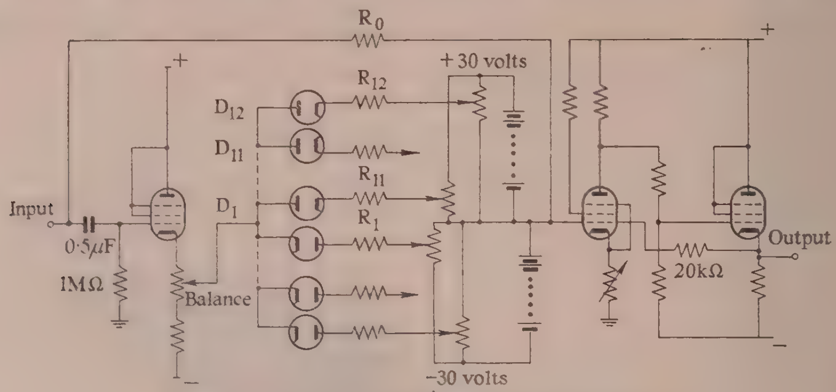


Fig. 4.—Circuit of the non-linear function generator.
$R_0 = 4 \cdot 7$ MΩ. $R_1 = 2$ MΩ. $R_{12} = 20$ kΩ.

order to minimize drift in the high zero-frequency gain elements of the linear system.

Departures from the ideal cubic characteristic occur for very large inputs when saturation occurs, and for very small inputs, for which the element behaves linearly.

(4) THE LINEAR SYSTEM

Fig. 5 shows the envelope of the theoretical describing function for the non-linear element employed, the heavy line representing

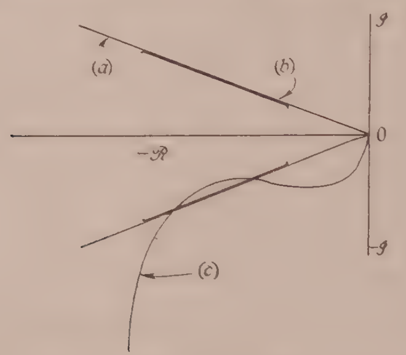


Fig. 5.—Derivation of the required frequency response.
(a) Envelope of theoretical describing function.
(b) Envelope of practical describing function.
(c) Required frequency-response locus of the linear system.

the region of proper operation of the unit. To ensure correct operation of the element where sub-harmonics are to be found, it is essential that the open-loop frequency-response locus of the linear system shall enter the critical area only where the practical unit functions correctly. Hence the required open-loop frequency response is of the form shown. It is desirable for the response to fall as the frequency is increased, so that the open-loop system must have the following characteristics:

(a) 90° phase shift at high frequencies.
(b) 90° phase shift at low frequencies.
(c) A phase shift of between 160° and 180° at some intermediate frequency.

The 90° phase change is provided by a high-gain Miller integrator, which operates over the whole of the frequency range considered. The further phase lag required is given by elements of the form shown in Fig. 6(a).

Using the notation shown in the Figure, the maximum phase shift θ_{max} is given by

$$\tan \theta_{max} = \frac{1}{2\sqrt{\left[\frac{R_2}{R_1}\left(1 + \frac{R_2}{R_1}\right)\right]}}$$

To give a phase change of 70° with one such element introduces a large attenuation at this frequency, and it is found preferable to use two elements in cascade, in which case

$$\frac{R_2}{R_1} = 0 \cdot 35$$

The frequency at which the phase shift is a maximum is given by

$$\omega^2 = 0 \cdot 386 C_2 R_2^2$$

The block diagram of the system is shown in Fig. 6(b).

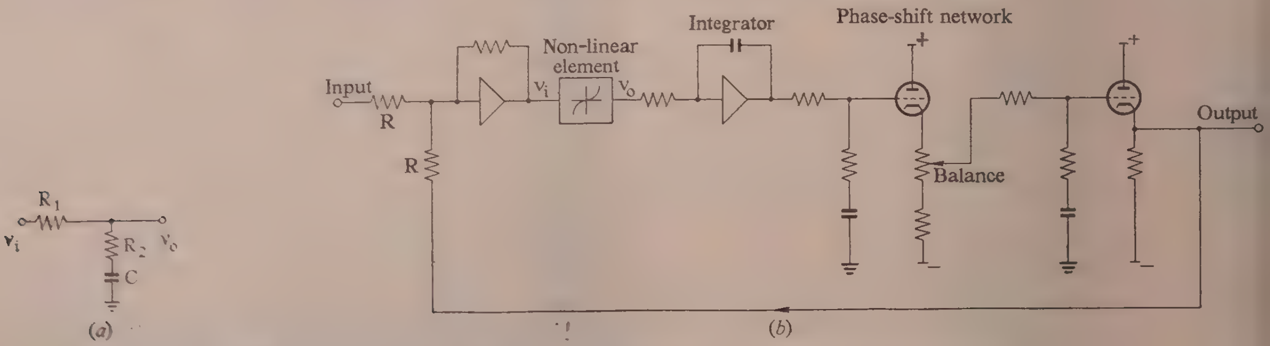


Fig. 6.—(a) A phase-shift network. (b) Block diagram of the experimental system.

(5) THE RESPONSE OF THE SYSTEM

The frequency response of the above system in the absence of ub-harmonics may be evaluated using known techniques.[7,8] This gives the variation of v_i with the amplitude and frequency of the input. The describing function for this value of v_i is hen considered. If this cuts the open-loop frequency-response ocus at a frequency of one-third the input frequency, continuous ub-harmonic oscillations can result. These oscillations can be considered as an instability of the closed loop, owing to the additional phase lag which is introduced by the non-linearity when a forcing sinusoidal signal is applied to the system.

From Fig. 5 it is evident that the frequency of these sub-armonic oscillations is approximately equal to the natural requency of the system, so that the fundamental output is very mall for an input of three times this frequency. Thus it is egitimate to neglect the fundamental output amplitude initially n comparison with the input amplitude.

It is to be expected that the sub-harmonic oscillation will not be self-starting in this system, since for small amplitudes of sub-harmonic component the describing function is along the negative real axis and does not cut the open-loop frequency-response locus. This is verified experimentally, for no sub-harmonics are observed unless started by applying an impulse to the system.

The results for the generation of sub-harmonics are given in Fig. 7. This shows the theoretical and experimental region of input amplitude and frequency in which sub-harmonics are generated. No stable sub-harmonic frequency of order other than one-third has been observed in the experimental system.

Fig. 8 shows the waveforms at the input and output of the system when sub-harmonic resonance occurs.

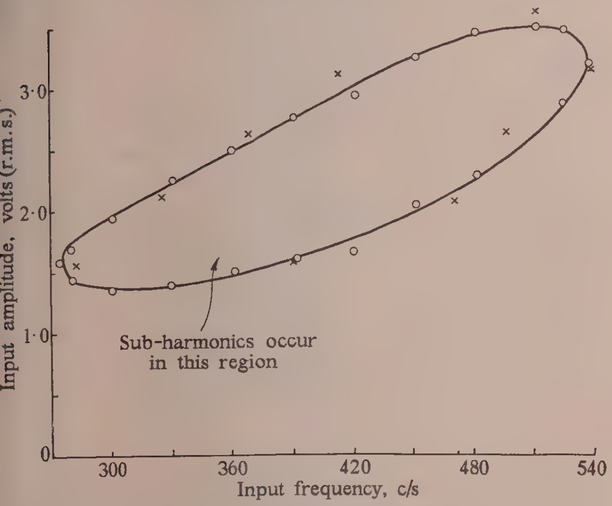
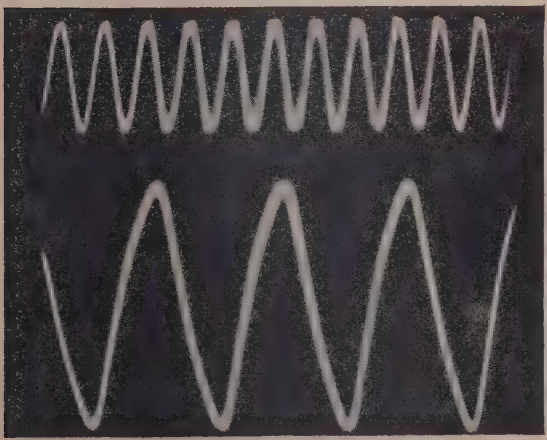


Fig. 7.—The region in which sub-harmonic resonance can occur.

⊙ Experimental points.
× Theoretical points.



(a)

(b)

Fig. 8.—Waveforms at the input and output of the system when sub-harmonic resonance occurs.

(a) Input. (b) Output.

(6) CONCLUSION

A method has been presented giving a graphical analysis of the generation of sub-harmonic frequencies. These sub-harmonics are shown to be due to the input signal modifying the response of the non-linear element to a further signal. This can result in a phase change for a further signal whose frequency is a sub-multiple of the input frequency.

In this manner a system which is expected to be stable for all inputs may produce troublesome forced oscillations. General criteria are developed showing the conditions for the absence of these sub-harmonic oscillations. Experimental work agrees with these results for a particular system.

Further work is proceeding on the analysis of systems where the non-linear element is directly coupled into the feedback loop. In this case, a second-order sub-harmonic may be generated, owing to a steady voltage at the input to the non-linearity making its response asymmetrical in form. In such cases, calculation of the describing function becomes extremely laborious.

(7) REFERENCES

(1) Hayashi, C.: "Subharmonic Oscillations in Non-Linear Systems," *American Journal of Applied Physics*, 1953, **24,** No. 5, p. 521.

(2) Pedersen, P. O.: "Subharmonics in Forced Oscillations in a Dissipative System" (Danmarks Naturvidenskabelige Samfund Publikationer, 1933).

(3) Stoker, J. J.: "Non-Linear Vibrations" (Interscience Publishers, New York, 1950).

(4) Kochenburger, R. J.: "A Frequency Response Method for Analyzing and Synthesizing Contactor Servomechanisms," *Transactions of the American I.E.E.*, 1950, **69,** p. 270.

(5) West, J. C.: "The Nyquist Criterion of Stability," *Electronic Engineering*, 1950, **22,** p. 169.

(6) Williams, F. C., and Ritson, F. J. U.: "Electronic Servo Simulators," *Journal I.E.E.*, 1947, **94,** Part IIA, p. 112.

(7) West, J. C., and Douce, J. L.: "The Frequency Response of a Certain Class of Non-Linear Feedback System," *British Journal of Applied Physics*, 1954, **5,** p. 204.

(8) West, J. C., and Nikiforuk, P. N.: "The Behaviour of a Remote Position Control Servomechanism with Hard-Spring Non-Linear Characteristics," *Proceedings I.E.E.*, Paper No. 1621 M, March, 1954 (**101,** Part II, p. 481).

(9) Johnson, E. C.: "Sinusoidal Analysis of Feedback-Control Systems containing Non-Linear Elements," Applications and Industry, *Transactions of the American I.E.E.*, 1952, **71,** p. 169.

(10) Nyquist, H.: "Regeneration Theory," *Bell System Technical Journal*, 1932, **11,** p. 126.

(8) APPENDIX

Any single-valued non-linearity may be expressed in the form

$$v_o = \sum_{m=0}^{m=\infty} A_m(v_i)^m \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

Let the input be

$$v_i = a\cos\omega t + b\cos(n\omega t + \phi) \quad . \quad . \quad . \quad (15)$$

where n is greater than unity, but is not necessarily an integer.

The output is of the form

$$\sum_{m=0}^{\infty}\sum_{r=0}^{m} B_r \cos^r \omega t \cos^{m-r}(n\omega t + \phi) \quad . \quad . \quad (16)$$

This may be written

$$v_o = \sum_{m=0}^{\infty}\sum_{r=0}^{m} B_r(\varepsilon^{j\omega t} + \varepsilon^{-j\omega t})^r[\varepsilon^{j(n\omega t+\phi)} + \varepsilon^{-j(n\omega t+\phi)}]^{m-r} \quad . \quad (17)$$

The bracketed terms may be expanded to give terms of the form

$$(\varepsilon^{j\omega t} + \varepsilon^{-j\omega t})^r = \sum_{p=0}^{r} C_p \varepsilon^{j(r-2p)\omega t}$$

$$[\varepsilon^{j(n\omega t+\phi)} + \varepsilon^{-j(n\omega t+\phi)}]^{m-r} = \sum_{q=0}^{q=m-r} D_q \varepsilon^{j(m-r-2q)(n\omega t+\phi)} \quad . \quad (18)$$

Thus, multiplying the two general terms, the resultant is

$$v_o = \sum_{m=0}^{m=\infty}\sum_{r=0}^{r=m}\Big[\sum_{p=0}^{r}\sum_{q=0}^{m-r} E_{p,q}\varepsilon^{j\{[(m-r-2q)n+r-2p]\omega t+(m-r-2q)\phi\}}\Big] \quad . \quad . \quad . \quad . \quad . \quad (19)$$

Thus for a resultant of angular frequency ω it is necessary for

$$(m - r - 2q)n + (r - 2p) = 1$$

i.e.

$$n = \frac{1 - r + 2p}{m - r - 2q} \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

where m, r, p and q are all integers

and

$$0 \leqslant r \leqslant m$$
$$0 \leqslant p \leqslant r$$
$$0 \leqslant q \leqslant m - r$$

It follows from eqns. (19) and (20) that a phase change can be introduced in all the intermodulation products of angular frequency ω, since ϕ appears in the output multiplied by the term $(m - r - 2q)$. This term is never zero for any finite value of n which satisfies eqn. (20). By substitution, the following are found to be the only possible values of n for the given order of m

$m = 1$. No solution.
$m = 2$. $n = 2$.
$m = 3$. $n = 3$.
$m = 4$. $n = 4, 3, 3/2$.
$m = 5$. $n = 5, 3, 2$.
$m = 6$. $n = 6, 4, 2, 3/2, 4/3$.
$m = 7$. $n = 7, 5, 3, 2, 5/3$.

It follows that, if the non-linearity contains only terms of order m_0 and less, the highest value of n is $n = m_0$. With such a non-linearity inserted in a feedback loop, the highest order of sub-harmonic oscillation possible has a frequency $1/m_0$ of the input frequency.

[The discussion on the above paper will be found on page 594.]

621.398 : 621–52

Paper No. 1728 M
Oct. 1954

THE RESPONSE OF REMOTE-POSITION-CONTROL SYSTEMS WITH HARD-SPRING NON-LINEAR CHARACTERISTICS TO STEP-FUNCTION AND RANDOM INPUTS

By JOHN C. WEST, B.Sc., Ph.D., Associate Member, and P. N. NIKIFORUK, B.Sc., Ph.D.

(*The paper was first received* 22*nd March, and in revised form* 1*st July*, 1954. *It was published in October*, 1954, *and was read before the* MEASUREMENTS SECTION 15*th March*, 1955.)

SUMMARY

The paper is concerned with the response to step-function and random inputs of a closed-loop system having a specific class of non-linearity called "hard-spring."

In order that an analysis be carried out for step-function inputs the system is restricted to that with a second-order characteristic equation. Two amplitude characteristics are considered, the first being a continuous curve, and the second being composed of segments of three straight lines. An experimental system is described, and the results are compared with analysis. Stabilization is achieved by velocity feedback, but derivative of error stabilization is also considered. A discussion of the effect of the non-linear characteristics on the overall stability of the system is included.

The response of the system with the straight-line characteristics is obtained for random inputs and in the presence of noise. Detailed results and a discussion of their implications and value compared with transient and frequency-response testing are given. Theoretical verification is not included.

LIST OF SYMBOLS

v = Instantaneous input signal to the non-linearity.
A = Amplitude of the input signal to the non-linearity.
$N(v)$ = Non-linear characteristic.
w = Instantaneous output from the non-linearity.
θ_i = Instantaneous input signal to the servomechanism.
θ_o = Instantaneous output signal from the servomechanism.
e_i = Normalized instantaneous input signal to the servomechanism.
e_o = Normalized instantaneous output signal from the servomechanism.
$T = 1/\omega_n$ = Time constant of the identical integrators.
T_d = Damping time-constant using velocity-feedback stabilization.
T_{d1} = Damping time-constant using phase-advance stabilization.
T_{d1c} = Time-constant for critical damping, using phase-advance stabilization.
G_1 = D.C. gain of velocity-feedback loop.
G = D.C. gain of the error-detecting unit.
α, β = Constants of the cubic non-linearity.
$v_d = \sqrt{(\alpha/\beta)}$ = An arbitrary constant.
$\psi = v/v_d$ = A non-dimensional variable.
$Y = \dot{e}_o/\omega_n$.
$X = e_o - e_i$.
S = Slope of a trajectory at any point on the phase plane.
V_c = Control signal level.
k, m, δ = Constants of segmented-line characteristics.
$A_\psi = A/v_d$.
E_i = (peak amplitude of input signal θ_i)/v_c.
D_e = equivalent damping of the non-linear system stabilized by velocity feedback.
D_{ec} = Critical damping for the linear system.
S_i = Normalized sinusoidal signal in the presence of noise.
N_{is} = Normalized random signal in the presence of noise.
N_{in} = Normalized noise signal.

Drs. West and Nikiforuk are in the Electrical Engineering Department, University of Manchester.

(1) INTRODUCTION

In many non-linear feedback systems, the nature of the non-linear characteristic determines some distinctive feature in the system behaviour. This is usually of an undesirable nature, but occasionally a particular type of non-linearity yields one feature which may find a field of application. Such an interesting characteristic is the "hard spring";[6] recent work[1] has shown that the introduction of a hard-spring non-linearity actually decreases the response time, apparently without affecting stability. In addition, it is found that the step response time continues to decrease with increasing step amplitude. It is also worth noting that it is thought[12] that the tensioning of muscle in animals (and man) follows a cubic relationship with the stimulating signal.

In linear analysis it is well known that the stabilization of a remote-position-control (r.p.c.) system by velocity feedback leads to a velocity lag, so that a persistent positional error exists when motion occurs. In many systems this is no disadvantage, but often derivative-of-error stabilization must be used to eliminate this velocity lag, and sometimes for other reasons.

These considerations still hold for similar r.p.c. systems having a hard-spring characteristic. For this reason work has been done on systems with both types of stabilization.

The non-linear control systems under consideration form a large class[1,2,3,4,5] whose characteristic equation can usefully and conveniently be altered to a second-order linear differential equation with constant coefficients when operating in the linear regime.

The response of non-linear servomechanisms is amplitude dependent and hence the input amplitude must be regarded as a parameter of the system. In investigating the performance of these systems the following parameters are necessary to specify the operating conditions:

(*a*) *Input amplitude*.—This is made non-dimensional by using the ratio of the input amplitude to some convenient constant amplitude in the non-linear characteristic.
(*b*) *Degree of non-linearity*.—This refers to the constants which define the non-linear characteristic.
(*c*) *Degree of damping*.—To be independent of (*a*) and (*b*) this refers to the equivalent linear system obtained when the non-linear element is removed and replaced by an equivalent linear element.

(2) A SERVOMECHANISM WITH A HARD-SPRING CHARACTERISTIC

The system to be considered, which is shown in Fig. 1, is an idealized second-order r.p.c. system with negligible friction, stabilized by output-velocity feedback and possessing a smooth hard-spring characteristic.

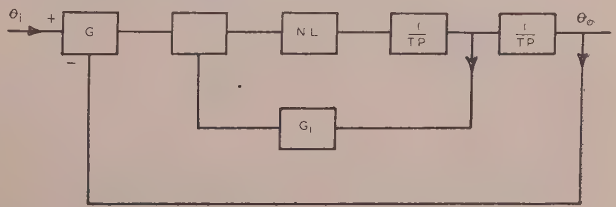


Fig. 1.—Block schematic of the system.

If the relationship between the input to and output from the non-linearity is of the form

$$w = N(v) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

then the equation of motion of the system is

$$T^2\ddot{\theta}_0 = N[G(\theta_i - \theta_o) - T_d\dot{\theta}_o] \quad . \quad . \quad . \quad . \quad (2)$$

where

$$T = 1/\omega_n \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

(2.1) The Cubic Characteristic

The non-linear characteristic is taken to be of the form

$$w = \alpha v + \beta v^3 \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where α and β are constants. The first parameter is dimensionless, whereas the second has the dimensions of v^{-2}. In a servomechanism the input signal is a measure of physical entity, and the slope of the characteristic is therefore determined with reference to a particular system using a particular physical quantity. It is preferable to stipulate a required characteristic without such reference.

This can be done by defining a non-dimensional variable ψ in terms of a particular value of the input quantity. Thus, let

$$v = \psi v_d \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where v_d is an arbitrarily chosen constant having the dimensions of v.

Eqn. (4) may be written in the form

$$w = v_d\alpha\left(\psi + \frac{\beta}{\alpha}v_d^2\psi^3\right) \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Since v_d is an arbitrary constant, any value which simplifies the above equation may be taken.

Let

$$v_d = \sqrt{(\alpha/\beta)} \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

Hence eqn. (6) becomes

$$w/v_d = \alpha(\psi + \psi^3) \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

(2.2) Degree of Damping

This refers to the equivalent linear system obtained when the cubic non-linearity is removed and is replaced by a linear amplifier of gain α; i.e. $\beta = 0$ in eqn. (4).

(2.3) Extension of Lienhard's Construction

For convenience it is generally arranged that the steady-state position determined by the final position of the input shall be at the origin. To obtain this, the system is assumed at rest at a point $-E_i$ until an input step of magnitude E_i is applied. For many second-order systems this response is the same as that obtained when the input is held at zero level and the output is released with an initial error. The resulting characteristic equation may be of the form

$$\ddot{\theta}_o + \mathrm{g}(\dot{\theta}_o) + \mathrm{f}(\theta_o) = 0 \quad . \quad . \quad . \quad . \quad . \quad (9)$$

where g and f may or may not be linear functions of velocity and position. The simplest general method of determining the trajectories of eqn. (9) is one first described by Lienhard.[6]

Of particular interest to servomechanism theory is the characteristic equation of the form

$$\ddot{\theta}_o + \mathrm{h}(\dot{\theta}_o, \theta_o) = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

where the function h is not separable into the two functions g and f of eqn. (9).

An extension of Lienhard's method[7] can be used to determine the trajectories of eqn. (10). If a control signal

$$\theta_c = -(\theta_o + T_d\dot{\theta}_o)$$

is fed into an amplifier whose amplitude characteristic is non-linear, the output can be of the form

$$-\mathrm{h}(\theta_o + T_d\dot{\theta}_o)$$

and need only be known graphically as a result of direct experimental tests on the amplifier. Eqn. (10) can be written in the form

$$\frac{d\dot{\theta}_o}{d\theta_o} = -\frac{\mathrm{h}(\theta_o + T_d\dot{\theta}_o)}{\dot{\theta}_o} \quad . \quad . \quad . \quad . \quad (11)$$

A trajectory is constructed by starting at the required point usually on the horizontal axis, and finding the gradient at this point. This is drawn as a short line. A second point is chosen at the tip of this short line and a second gradient is obtained. Thus the entire path is constructed of small segments. The method of obtaining the gradient at a specific point is as follows:

(*a*) The axes of the phase-plane must be graduated in similar scales.

(*b*) Let the co-ordinates of the point P, Fig. 2, for which the gradient is required be θ_{o1} and $\dot{\theta}_{o1}$. The output velocity $\dot{\theta}_{o1}$ is

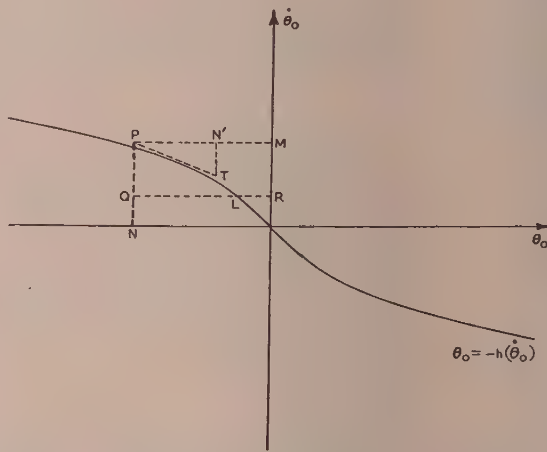


Fig. 2.—Graphical extension of Lienhard's construction.

multiplied by T_d, which can be done graphically, and is added algebraically to θ_{o1}. A vertical line NQ is drawn through P where

$$NQ = \theta_{o1} + T_d\dot{\theta}_{o1}$$

(*c*) A graph of the non-linear characteristic of the amplifier in the form

$$\theta_o = -\mathrm{h}(\dot{\theta}_o)$$

is drawn on the same figure.

(*d*) A horizontal line is drawn through Q intersecting the axis at R and the function

$$\theta_o = -\mathrm{h}(\dot{\theta}_o) \text{ at L}$$

Then LR represents the function $-\mathrm{h}(\theta_o + T_d\dot{\theta}_o)$ at $\theta_o = \theta_{o1}$, $\dot{\theta}_o = \dot{\theta}_{o1}$.

(*e*) A point N′ is marked off on PM such that $PN = PN' = \dot{\theta}_{o1}$. A normal N′T and of magnitude LR is erected on PM at N′, the direction of the normal being such that it points in the direction of the positive $\dot{\theta}_o$ axis if $-\mathrm{h}(\theta_{o1} + T_d\dot{\theta}_{o1})$ is positive, and in the direction of the negative $\dot{\theta}_o$ axis if $-\mathrm{h}(\theta_{o1} + T_d\dot{\theta}_{o1})$ is negative. The required gradient at P is given by an extension of the line PT (note direction).

(*f*) This construction is repeated for a neighbouring point P_1.

(2.4) The Phase-Plane Portrayal of the Step-Response

Without the direct solution of eqn. (11) it is possible to btain a qualitative idea of the step-function reponse of a imulator possessing a cubic non-linearity of the form given by qn. (10) using the above extension of Leinhard's construction.

As developed in Appendix 11.1 the normalized equation of notion of the simulator of Fig. 1 possessing a cubic non-linear haracteristic is

$$\frac{dY}{dX} = -\alpha\frac{[(GX + T_d\omega_n Y) + (GX + T_d\omega_n Y)^3]}{Y} \quad . \quad (12)$$

he slope S of a trajectory at any point on the phase-plane being qual to dY/dX.

There are four cases of particular interest.

a) $S = \infty$: The trajectory is vertical and the locus is $Y = 0$.
b) $S = 0$: There is only one real solution of eqn. (12).

The locus that passes through points of zero slope on the rajectories is a straight line through the origin whose equation is

$$Y = -\frac{G}{T_d\omega_n}X \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

(c) Variation of the trajectory slope on the Y axis. For this ase eqn. (12) reduces to

$$S = -\alpha T_d\omega_n[1 + (T_d\omega_n Y)^2] \quad . \quad . \quad . \quad . \quad (14)$$

he slope increases rapidly as Y is increased.

(d) $S = -2G/T_d\omega_n$ for the equivalent linear simulator. For his condition the linear simulator is critically damped and the quation of the straight-line trajectory in the phase-plane is etermined from eqn. (43) and (44) to be

$$Y = -\frac{2G}{T_d\omega_n}X \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

'or this damping there is a limiting boundary for the non-linear imulator, which cannot be crossed by any trajectory. This oundary passes through the origin. The system is approxi- nately linear for small values of X and Y, and therefore near the rigin the slope of the limiting boundary will be equal to that f the straight-line trajectory of the linear system. Starting at he origin, the required trajectory can be drawn by the con- truction described in Section 2.3. The limiting boundary and wo non-linear trajectories are shown in Fig. 3. In addition, wo trajectories of the equivalent linear simulator are shown in roken line. Two non-linear trajectories and the two corre- ponding linear trajectories are shown in Fig. 4 for small damping.

2.5) Pattern of the Response as Displayed on the Phase-Plane

If the linear simulator is critically damped there is no over- hoot for either system. The rise time of the non-linear simu- ator is smaller, and this is indicated by the output velocity of the on-linear system being greater than that of the linear system hroughout the entire response. The ratio of the maximum elocity attained by the non-linear simulator to that attained by he linear simulator for the same step input increases with an ncrease of the input amplitude. Consequently, the ratio of the orresponding rise times decreases.

The same discussion regarding the rise time applies to the nderdamped condition. Typical linear and non-linear tra- ectories are shown in Fig. 4. The velocity of the non-linear ystem is considerably greater than that of the linear system hroughout most of the response, becoming smaller only near the rst overshoot. The acceleration and deceleration of the non- near system are greater than those of the linear system.

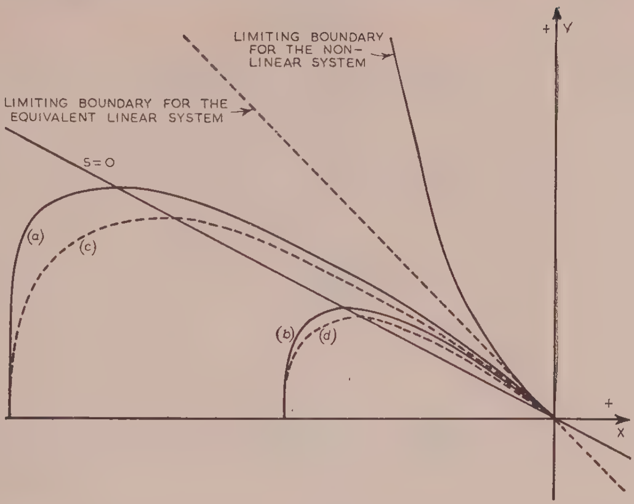


Fig. 3.—Step-function response in the phase plane for a critically damped system with a cubic non-linearity.

(*a*), (*b*) Non-linear responses.
(*c*), (*d*) Corresponding linear responses.

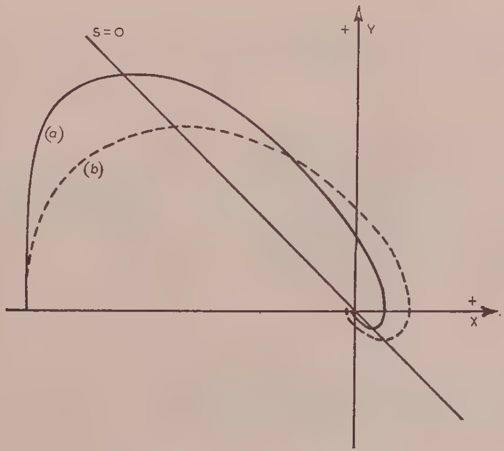


Fig. 4.—Step-function response in the phase plane for a lightly damped system with a cubic non-linearity.

(*a*) Non-linear response.
(*b*) Corresponding linear response.

For underdamped systems the ratio of the first overshoot of the non-linear system to that of the linear system for the same step input can be taken as a measure of the improvement in the transient response.

(3) A SECOND-ORDER SIMULATOR POSSESSING A SEGMENTED-LINE CHARACTERISTIC

Physically straight-line segmented characteristics can be achieved more accurately and more easily than smooth charac- teristics of the form given in eqn. (4) and are therefore more adaptable to practical systems. It is in this respect that an analysis has been carried out for a segmented-line characteristic.

(3.1) The Straight-Line Segmented Characteristic

The experimental investigation has been limited to that particular characteristic[1] shown in Fig. 5. It consists of three

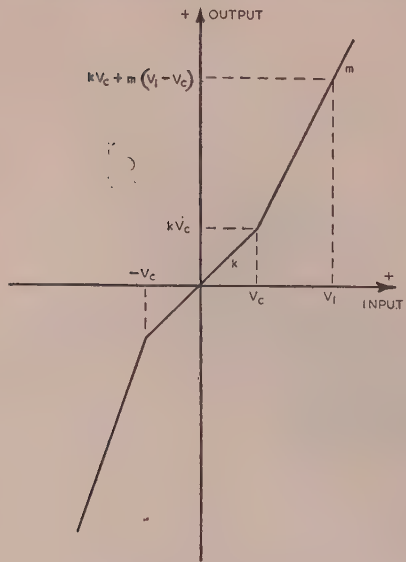


Fig. 5.—The segmented-line non-linear characteristic.

$V_c = 1 \cdot 5$ volts.
$k = 0 \cdot 94$.

straight lines, and non-linear operation occurs sharply at the control signal level $\pm V_c$.

(3.2) The Non-Linear Element

The non-linear element[1] consists of two amplifiers in parallel with a common input and their outputs added together. One amplifier has a small constant gain k over the whole operating range. The other has a constant gain M_a that can be varied; by a diode arrangement the amplifier is made operable only when the input magnitude is greater than $\pm V_c$. Hence

$$m = k + m_a$$

(3.3) Degree of Damping

This refers to the equivalent linear system obtained when the non-linearity is replaced by a linear amplifier of gain k.

(3.4) The Analogue Computer

The computer used in this work simulated an elementary second-order r.p.c. system with the inclusion of the segmented-line characteristic and stabilized by output-velocity feedback.

The design of this simulator is conventional[8,9] and little detail will be given here. A block schematic is shown in Fig. 1, where all the elements incorporate negative-feedback amplifiers of the anode-follower type. The integrators are of the Miller type.

(4) ANALYSIS OF THE STRAIGHT-LINE SEGMENTED SYSTEM

There are three conditions under which a simulator possessing the non-linear characteristic of Fig. 5 operates:

Condition 1 $\quad |V_p| < V_c$
Condition 2 $\quad V_p > V_c$
Condition 3 $\quad V_p < - V_c$

where V_p is the peak value of the control signal

$$G(\theta_i - \theta_o) - T_d\dot{\theta}_o$$

The normalized equations of motion (see Appendix 11.2) fo the three conditions are

$$Y\frac{dY}{dX} + \omega_n T_d k Y + kGX = 0 \text{ for condition 1} \quad . \quad (16)$$

$$Y\frac{dY}{dX} + \omega_n T_d k(m/k) Y + mGX = -\delta \text{ for condition 2} \quad . \quad (17)$$

$$\text{and } Y\frac{dY}{dX} + \omega_n T_d k(m/k) Y + mGX = \delta \text{ for condition 3} \quad . \quad (18)$$

(4.1) The Phase-Plane for the Straight-Line Segmented Characteristic

Owing to the sharp change from linear to non-linear operatio it is possible to draw boundaries on the phase-plane correspond ing to the limits of each mode of operation. As a result, th phase-plane can be divided into three regions, each havin characteristic trajectories.

The boundaries between linear and non-linear modes o operation are given by

$$Y = -\frac{1}{T_d\omega_n}(GX \pm 1) \quad . \quad . \quad . \quad . \quad (19)$$

and are shown in Fig. 6.

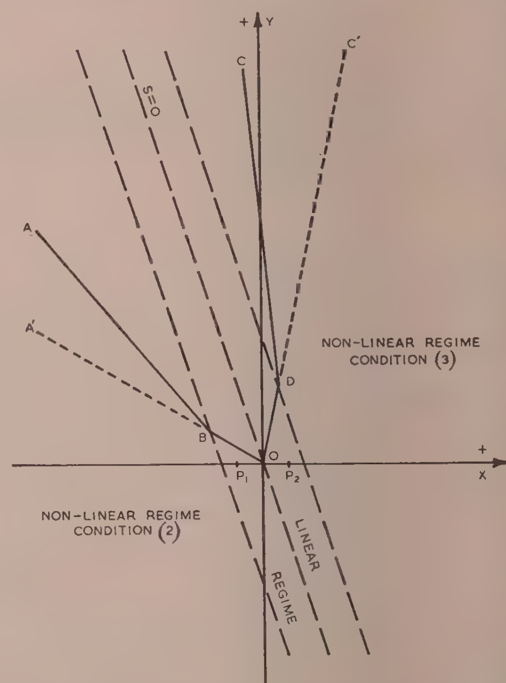


Fig. 6.—Loci of constant slope and the boundaries dividing the phase-plane into regions of linear and non-linear operation.

$P_1 = -\frac{\delta}{mG}, 0. \quad P_2 = \frac{\delta}{mG}, 0$

It is possible to obtain a qualitative idea of the trajectories in the three regimes without solving the normalized equations of motion. The types of trajectories obtained in the linear region are well known[10,11] and will be dealt with very briefly.

The locus of all points in the linear regime having a given slope S is

$$Y = -\frac{kGX}{S + \omega_n T_d k} \quad . \quad . \quad . \quad . \quad (20)$$

There are four cases of particular interest:

(a) $S = \infty$. The trajectory is vertical and the locus is $Y = 0$.

(b) $S = 0$. The trajectory is horizontal and the locus is

$$Y = -\frac{GX}{\omega_n T_d} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

(c) $S = -\omega_n T_d k$. The locus is $X = 0$.

(d) A straight-line trajectory for critical damping. The locus is found from eqn. (16) to be

$$Y = -\frac{\omega_n T_d k}{2} X \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

For the condition $V_p > V_c$ the locus of all points having a given slope S is

$$Y = -\frac{mGX + \delta}{S + (\omega_n T_d k)(m/k)} \quad . \quad . \quad . \quad . \quad (23)$$

and this equation defines a family of straight lines passing through the point

$$X = -\frac{\delta}{mG}, \quad Y = 0$$

For the condition $V_p < -V_c$ the locus is

$$Y = -\frac{mGX - \delta}{S + (\omega_n T_d k)(m/k)} \quad . \quad . \quad . \quad . \quad (24)$$

and this equation defines a family of straight lines passing through the point

$$X = \frac{\delta}{mG}, \quad Y = 0$$

For the non-linear mode of operation there are four cases of particular interest.

(a) $S = \infty$.
The trajectory is vertical and the locus is $Y = 0$.

(b) $S = 0$.
For condition 2 the locus is found from eqn. (23) to be

$$Y = -\frac{1}{T_d \omega_n}[GX + (1 - k/m)] \quad . \quad . \quad . \quad (25)$$

and for condition 3 the locus is found from eqn. (24) to be

$$Y = -\frac{1}{T_d \omega_n}[GX - (1 - k/m)] \quad . \quad . \quad . \quad (26)$$

The two conditions expressed by eqns. (25) and (26) do not hold since the loci of zero slopes fall completely within the region of linear operation. Therefore there is only one real solution for $S = 0$ expressed by eqn. (21).

(c) *The variation of the slope along the Y-axis.*
The variation of the slope S along the positive Y-axis can be found by putting $X = 0$ in eqn. (18).

Hence $$S = \frac{\delta}{Y} - \omega_n T_d k(m/k) \quad . \quad . \quad . \quad . \quad (27)$$

The slope therefore varies from a value of

$$S = -\omega_n T_d k$$

at the point of intersection of the positive Y-axis and the boundary

$$Y = -\frac{1}{T_d \omega_n}(GX - 1)$$

to a value $$S = -\omega_n T_d k(m/k)$$

at $Y = +\infty$.

Similarly it can be shown that along the negative Y-axis the slope varies from a value of

$$S = -\omega_n T_d k$$

at the point of intersection of the negative Y-axis and the boundary

$$Y = -\frac{1}{T_d \omega_n}(GX + 1)$$

to a value of $$S = -\omega_n T_d k(m/k)$$

at $Y = -\infty$.

(d) *The straight-line trajectories.*
If the simulator is critically damped in the linear region it is overdamped in the non-linear region. There are then (see Appendix 11.3) two straight-line trajectories terminating on the boundary

$$Y = -\frac{1}{T_d \omega_n}(GX - 1)$$

and another two straight-line trajectories of the same slope as the first two, terminating on the boundary

$$Y = -\frac{1}{T_d \omega_n}(GX + 1)$$

If the trajectories were extended into the linear regime the first two would pass through the point

$$X = \frac{\delta}{mG}, \quad Y = 0$$

and the latter two through the point

$$X = -\frac{\delta}{mG}, \quad Y = 0$$

Two of the trajectories have the slope

$$S_1 = \frac{-\omega_n T_d k(m/k) - \sqrt{\{[\omega_n T_d k(m/k)]^2 - 4mG\}}}{2} \quad . \quad (28)$$

and the other two have the slope

$$S_2 = \frac{-\omega_n T_d k(m/k) + \sqrt{\{[\omega_n T_d k(m/k)]^2 - 4mG\}}}{2} \quad . \quad (29)$$

It is shown in Appendix 11.4 that the line which is the locus of all points having a given slope S' in the linear regime intersects a boundary at the same point as the line which is the locus of all points having the same slope S' in the non-linear regime. The typical loci are shown in Fig. 6.

(4.2) The Step-Function Response in the Phase-Plane

The response for the following three values of damping is of particular interest:

(a) $D_e < D_{ec}\sqrt{(k/m)}$.

(b) $D_e = D_{ec}\sqrt{(k/m)} =$ (critical damping in the non-linear region).

(c) $D_e = D_{ec} =$ (critical damping in the linear region).

(4.2.1) The Step-Function Response for $D_e < D_{ec}/\sqrt{(k/m)}$.

For this damping the response of the system is underdamped in both modes of operation. The response can be studied with

the aid of Fig. 6. ABO and CDO are typical non-linear loci of constant slope, and A′BO and C′DO are the loci of the equivalent linear simulator. The effect of the non-linearity is to produce greater acceleration and deceleration than would be obtained with the equivalent linear simulator.

The result of this is shown in Fig. 7. A typical non-linear trajectory and the corresponding linear trajectory are shown in this Figure.

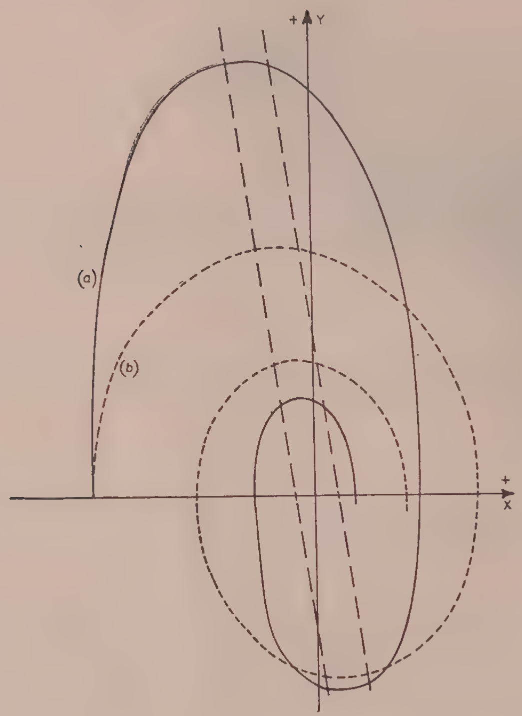


Fig. 7.—Step-function response in the phase plane of a system with a segmented-line characteristic.

$m = 5\cdot 0.$
$D_e = 0\cdot 1 D_{ec}.$

(*a*) Non-linear response.
(*b*) Corresponding linear response.

The overshoots of the non-linear simulator are much smaller than the corresponding overshoot of the linear simulator. The maximum velocity attained by the non-linear simulator, is much greater than that attained by the equivalent linear simulator, and the rise time of the non-linear simulator is much less than that of the linear simulator.

Fig. 8 shows the improvement in the first overshoot with increasing amplitude of the step input, and also with increasing non-linearity.

The ratio

$$\frac{\text{First overshoot of the non-linear simulator}}{\text{First overshoot of the equivalent linear simulator for the same step input}}$$

tends to a limit value which is given by the ratio

$$\frac{\text{First overshoot of the equivalent linear system of gain } M}{\text{First overshoot of the equivalent linear system of gain } k}$$

There is a similar improvement in the rise time.

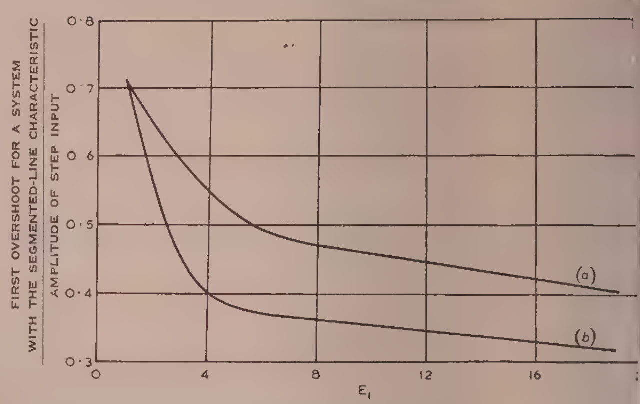


Fig. 8.—Improvement in the transient response of a system with a segmented-line characteristic.

(*a*) $m = 5\cdot 0.$
(*b*) $m = 15\cdot 0.$

(4.2.2) **The Step-Function Response for $D_e = D_{ec}\sqrt{(k/m)}$.**

For this damping the response in the non-linear regime is critically damped. There are then two straight-line trajectories of slope

$$S = -\frac{\omega_n T_d k(m/k)}{2}$$

and they are shown in Fig. 9.

The motion for a step input of $X = -\infty$ takes place along the straight-line trajectory AB. At B this trajectory intersects the boundary separating linear and non-linear modes of opera-

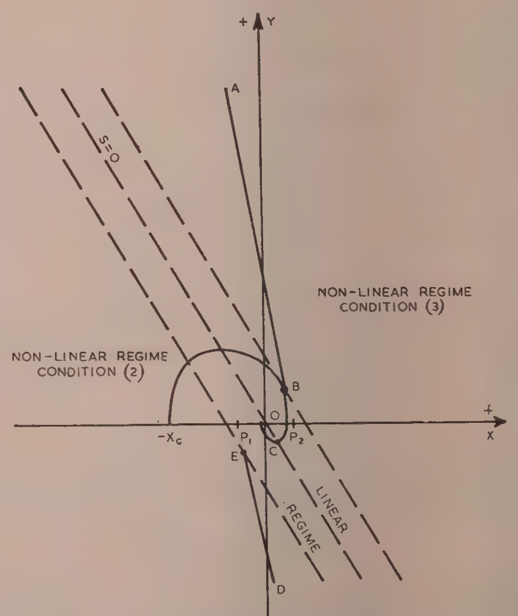


Fig. 9.—Limiting boundaries on the phase plane for a value of damping that is critical in the non-linear regime.

$P_1 = -\dfrac{\delta}{mG}, 0.$

$P_2 = \dfrac{\delta}{mG}, 0.$

tion. The subsequent motion in the linear region takes place along the trajectory BCO.

There is a certain minimum value of input $-X_c$ for which the motion after it has left B takes place along the trajectory BCO. The motion in this region, and consequently the magnitude of the first overshoot, is independent of the input for all inputs greater than $-X_c$. The input $-X_c$ may be determined by the extension described in Section 2.3.

The motion for a step input of $X = +\infty$ takes place along the straight-line trajectory DE.

A typical non-linear trajectory and the corresponding linear trajectory are shown in Fig. 10.

For this value of damping there is a very marked improvement in the transient response.

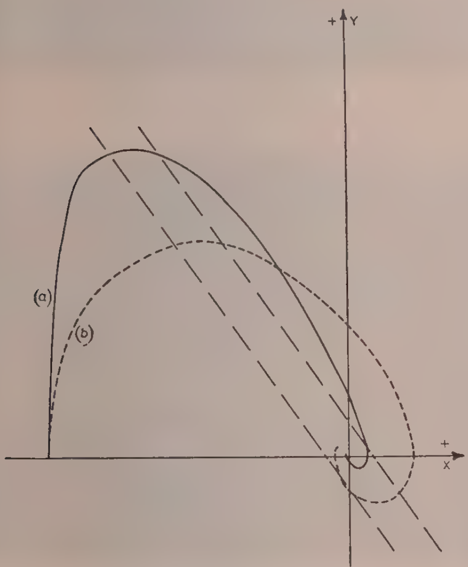


Fig. 10.—Step-function response in the phase plane of a system with a segmented-line characteristic.

$m = 5{\cdot}0.$
D_e is critical in the non-linear regime.
(a) Non-linear response.
(b) Corresponding linear response.

(4.2.3) The Step-Function for $D_e = D_{ec}$.

For this damping the response of the simulator is critically damped in the linear regime. There is one straight-line trajectory in the linear region of slope

$$S = -\frac{\omega_n T_d k}{2}$$

and four straight-line trajectories in the non-linear regime; two of the trajectories have the slope S_1 of eqn. (28), and the remaining two trajectories have the slope S_2 of eqn. (29).

If the input is released from rest at $X = -\infty$ the resulting motion is shown in the phase plane of Fig. 11 and occurs along the straight-line trajectory AB whose slope is S_2. At B the straight-line trajectory enters the linear regime and the subsequent motion approaches the straight-line trajectory CO of the linear regime. The limiting boundary consists, therefore, of the straight line AB and the curve BO.

For this value of damping, there is also a trajectory which passes through point B but which remains within the linear regime at B. This trajectory $-X_{c1}$BO can also be determined by the extension described in Section 2.3. For all values of X

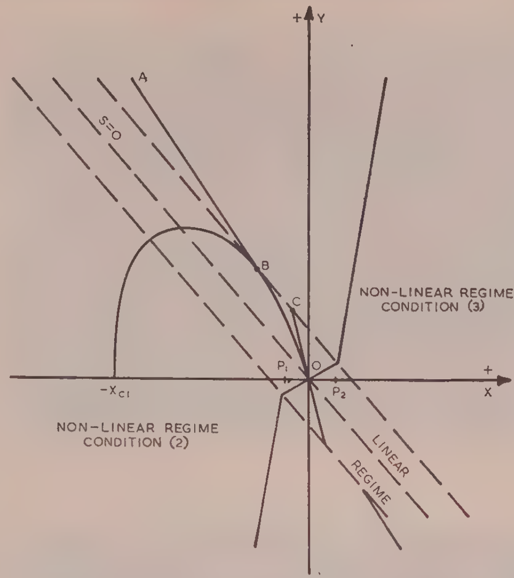


Fig. 11.—Limiting boundaries on the phase plane for a value of damping that is critical in the linear regime.

$$P_1 = -\frac{\delta}{mG}, 0.$$
$$P_2 = \frac{\delta}{mG}, 0.$$

numerically less than $-X_{c1}$ the response from point B to the origin occurs along curved portion BO of the limiting boundary.

A typical non-linear response is shown in Fig. 12. For this degree of damping there are no overshoots and the improvement in the transient response is small.

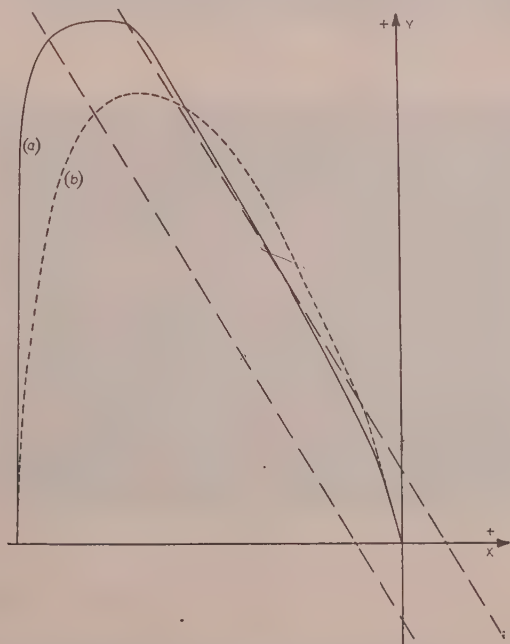


Fig. 12.—Step-function response in the phase plane of a system with a non-linear segmented-line characteristic.

$m = 5{\cdot}0.$
D_e is critical in the linear regime.
(a) Non-linear response.
(b) Corresponding linear response.

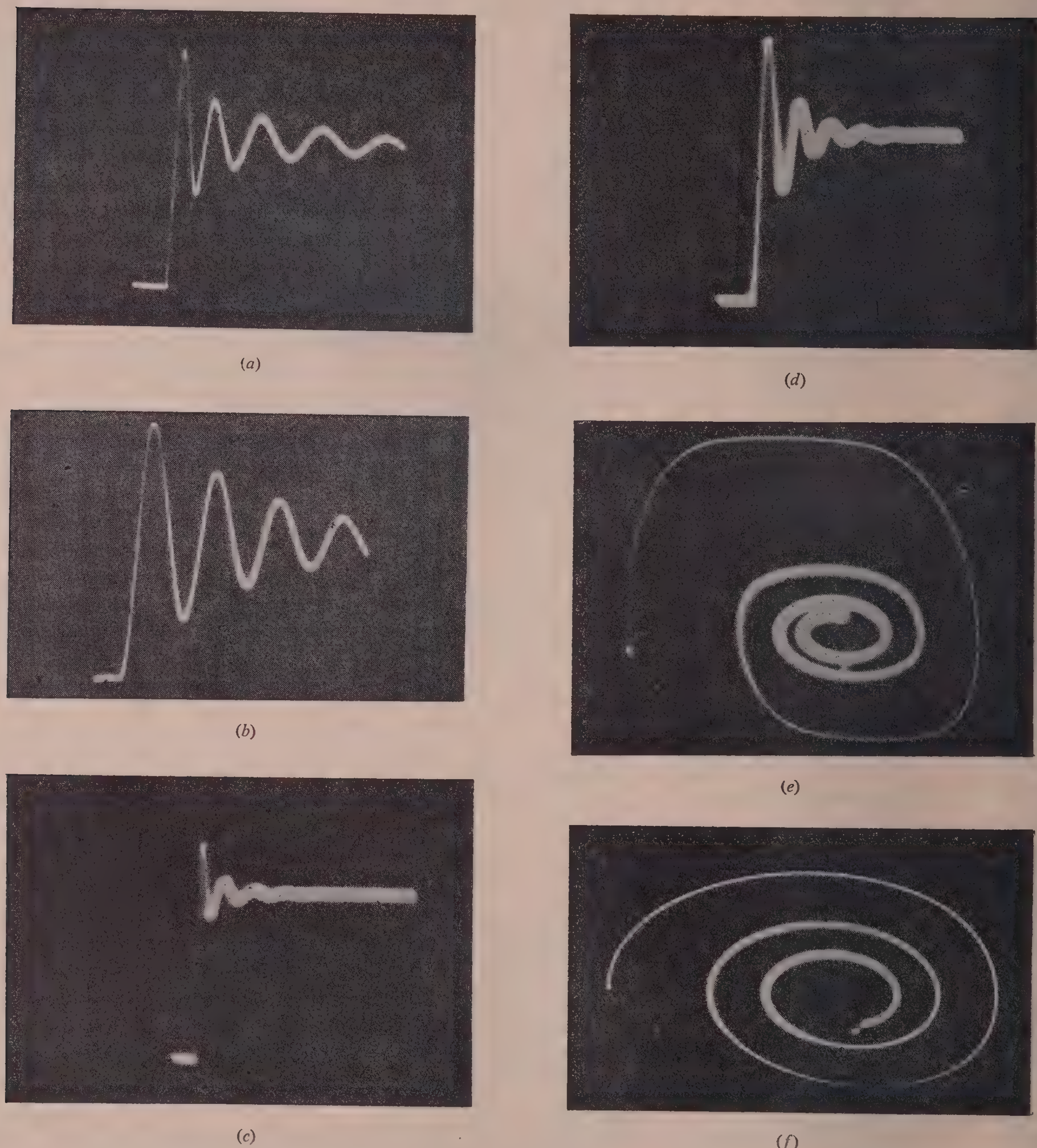


Fig. 13.—Experimental transient responses and phase-plane trajectories.

$m = 5\cdot0$.
$D_e = 0\cdot1D_{ec}$.

(*a*) Non-linear response, $E_i = 4\cdot3$.
(*b*) Corresponding linear response, $E_i = 4\cdot3$.
$D_e = 0\cdot2D_{ec}$.
(*c*) Non-linear response, $E_i = 10\cdot4$.

(*d*) Corresponding linear response, $E_i = 10\cdot4$.
$D_e = 0\cdot1D_{ec}$.
(*e*) Non-linear trajectory.
(*f*) Corresponding linear trajectory.

(5) EXPERIMENTAL RESULTS USING THE SEGMENTED-LINE CHARACTERISTIC

The experimental system was of the form shown in Fig. 1, and the non-linear characteristic was of the form shown in Fig. 5. V_c was made equal to $1 \cdot 5$ volts, k equal to $0 \cdot 94$ and m equal to $5 \cdot 0$. The amplitude of the step input was varied over an appropriate range. The following three values of damping were used:

$$0 \cdot 1\ D_{ec}.$$
$$D_{ec}\sqrt{(k/m)}.$$
$$D_{ec}.$$

The experimental results are shown in Figs. 13, 14, and 15.

The behaviour of the non-linear system for a value of damping equal to $0 \cdot 1 D_{ec}$ and for various values of input amplitude is shown in Fig. 13. The corresponding linear behaviour is also shown in this Figure. The response of the non-linear system to a small input is shown in Fig. 13(*a*). Comparing this with the corresponding linear response shown in Fig. 13(*b*) it is seen that there is small improvement.

The non-linear time response to a much larger input is shown in Fig. 13(*c*), and the corresponding linear response in Fig. 13(*d*). There is a marked improvement in the non-linear response. This experimental behaviour agrees favourably with the behaviour predicted from graphical analysis.

The non-linear response in the phase plane is shown in Fig. 13(*e*), and the corresponding linear response in Fig. 13(*f*). The non-linear response shows pronounced flattening at maximum and minimum velocity. This is due to the linear regime. The non-linear response also shows a marked increase in output velocity and a considerable reduction in the first overshoot.

The behaviour of the non-linear system for a value of damping $D_e = D_{ec}\sqrt{(k/m)}$ is shown in Fig. 14. The response of the non-linear system to a small input is shown in Fig. 14(*a*), and the response to a large input is shown in Fig. 14(*c*). The corresponding linear responses are shown in Figs. 14(*b*) and 14(*d*) respectively. Figs. 14(*a*) and 14(*c*) show that the overshoot is independent of the input amplitude.

The response in the phase plane of the non-linear system is shown in Fig. 14(*e*) for two values of input amplitude. The response near the origin is independent of the input amplitude, and this agrees favourably with the curves of Fig. 10. The corresponding linear response is shown in Fig. 14(*f*).

The reduction in the first overshoot of the non-linear response is considerable when the input signal is large. There is also an improvement in the rise time.

The non-linear response for a value of damping D_{ec} is shown in Fig. 15(*a*), and the corresponding linear response in Fig. 15(*b*). There is almost no difference in the responses. The response in the phase plane of the non-linear system is shown in Fig. 15(*c*), and the corresponding linear response in Fig. 15(*d*). These responses agree favourably with the curves shown in Fig. 12.

In Fig. 15(*c*) the non-linear response to the small input corresponds approximately to the trajectory obtained when the input is $-X_{c1}$ (see Section 4.2.3).

(6) DERIVATIVE-OF-ERROR STABILIZATION

If a step-function of amplitude h is applied to a phase-advance element whose characteristic equation is

$$1 + pT_{d1}$$

the output consists of a step of amplitude h and a spike of infinite amplitude and zero time duration. Consequently the response in the phase plane of a system stabilized by phase advance is different from the corresponding response of a system stabilized by velocity feedback.

(6.1) Basic Analysis

The closed-loop equation of an idealized linear second-order system stabilized by phase advance is

$$e_0 = \frac{kG(1 + pT_{d1})e_i}{p^2T^2 + pkGT_{d1} + kG} \quad . \quad . \quad . \quad . \quad (30)$$

where T_{d1} is the damping time-constant.

The effect of the spike upon the step-function response may be seen by differentiating eqn. (30) with respect to time.

Hence $$\dot{e}_0 = \frac{pkG(1 + pT_{d1})e_i}{p^2T^2 + pkGT_{d1} + kG} \quad . \quad . \quad . \quad . \quad (31)$$

Let a step-function $e_i = -h$ be applied at the time $t = 0$.

The limit of $\dot{e}_0$ at $t = \xi$, where ξ_1 is positive and approaching zero, is given by

$$\lim_{p\to\infty}\left[\frac{-pkG(1 + pT_{d1})h}{p^2T^2 + pkGT_{d1} + kG}\right]$$

This equals $$-kGT_{d1}h/T^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

The immediate effect of this inherent characteristic in a phase-advance element is to produce an instantaneous jump in the output velocity. The analysis of the subsequent motion in the phase plane is similar to that described in Section 4.1. The effect of this instantaneous jump upon the subsequent motion may be deduced by noting that as a result of it, the output velocity is greater for the same input than that of a system stabilized by output-velocity feedback. Consequently the rise time of the former is less than that of the latter, the overshoot of the former being greater than that of the latter.

(6.2) Experimental Results

Fig. 16(*a*) is the transient response of the non-linear simulator stabilized by phase advance, and Fig. 16(*b*) is the corresponding response of the linear system which is also stabilized by phase advance. Fig. 16(*c*) is the corresponding non-linear response of the system stabilized by velocity feedback; Figs. 17(*a*), 17(*b*), 17(*c*) and 17(*d*) are the corresponding phase-plane responses.

By comparing Fig. 16(*a*) with Fig. 16(*c*) it is seen that for the same damping, non-linearity and input amplitude the rise time of the system stabilized by phase advance is less than that of the system stabilized by velocity feedback; the first overshoot of the former is greater than the overshoot of the latter.

By comparing Fig. 17(*a*) with Fig. 17(*c*) it is seen that the output velocity of the system stabilized by phase advance is greater than that of the system stabilized by velocity feedback.

As shown in Figs. 17(*a*) and 17(*b*) there is a very rapid but not instantaneous increase in the velocity of a system stabilized by phase advance. This is because (*a*) there is always present an unwanted time-constant in the phase-advance unit and as a result the spike is of finite duration; and (*b*) saturation occurs when the amplitude of the spike is large.

It is exceedingly tedious to carry out an exact analysis of the transient response of a system stabilized by phase advance because of these two physical limitations.

(7) STABILITY CONSIDERATIONS

The open-loop equation of the non-linear system using derivative-of-error stabilization can be written in the form

$$p^2T^2e_o = N\left[G(e_i - e_o)\left(\frac{1 + pT_{d1}}{1 + p\xi T_{d1}}\right)\right]$$

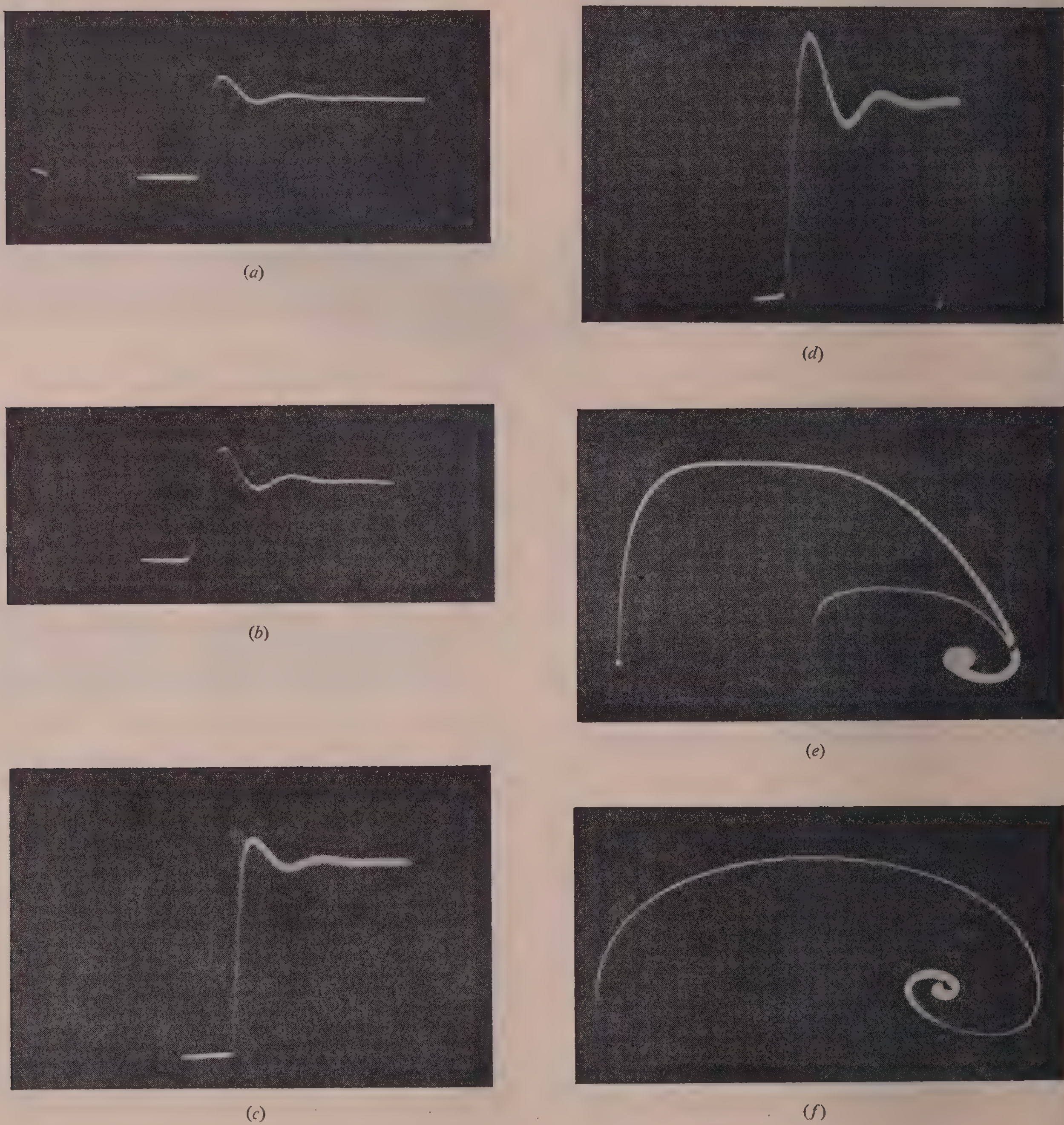


Fig. 14.—Experimental transient responses and phase-plane trajectories.

$m = 5 \cdot 0$.
D_e is critical in the non-linear regime.

(a) Non-linear response, $E_i = 4 \cdot 5$.
(b) Corresponding linear response, $E_i = 4 \cdot 5$.
(c) Non-linear response. $E_i = 10 \cdot 4$.
(d) Corresponding linear reponse, $E_i = 10 \cdot 4$.
(e) Non-linear trajectories.
(f) Corresponding linear trajectory.

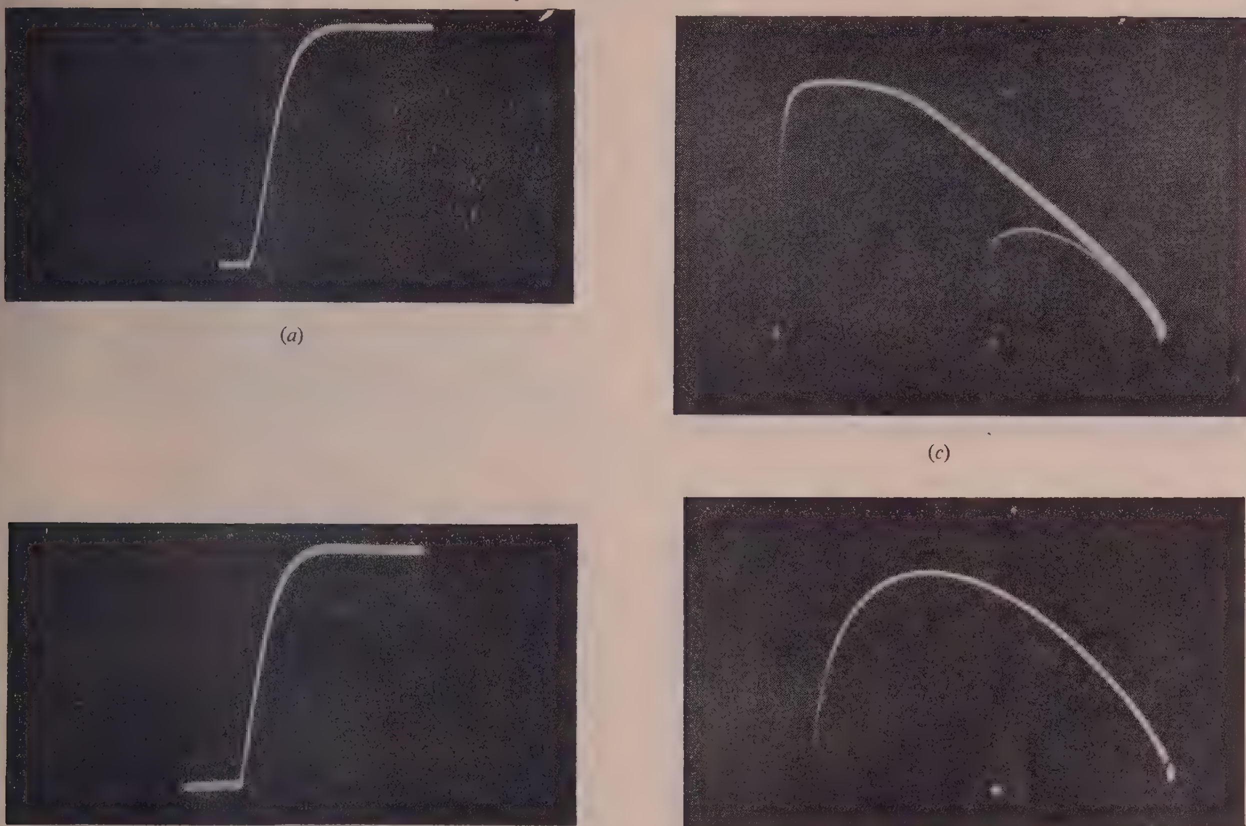


Fig. 15.—Experimental transient responses and phase-plane trajectories.

$m = 5 \cdot 0$.
D_e is critical in the linear regime.

(a) Non-linear response, $E_i = 10 \cdot 4$.
(b) Corresponding linear response, $E_i = 10 \cdot 4$.
(c) Non-linear trajectories.
(d) Corresponding linear trajectory.

where ξT_{d1} is the unwanted time-constant of the phase-advance unit.

When the input is sinusoidal useful information for frequency/response analysis and stability determination of non-linear systems can be obtained by resolving the output of the non-linear element into Fourier components. This is the "describing-function technique."[13] In most control systems only the fundamental component is of interest, the higher-order components being attenuated by the servo-motor.

(7.1) The Stability of a System possessing the Cubic Non-Linear Characteristic

If a sinusoidal input signal of frequency $\omega/2\pi$ is applied to the system and it is assumed that the harmonics in the output are negligible—owing to the attenuation by the integrators—then the error signal is also sinusoidal and is of the same frequency.

If the input signal to the cubic non-linearity is

$$v = A \sin \omega t \quad (33)$$

then from eqn. (5)

$$\psi = \frac{A}{v_d} \sin \omega t$$
$$= A_\psi \sin \omega t \quad (34)$$

The output is then

$$w/v_d = \alpha(A_\psi \sin \omega t + A_\psi^3 \sin^3 \omega t)$$

and

$$\frac{1}{\alpha}\frac{w}{v_d} = \left[A_\psi \sin \omega t + \frac{A_\psi^3}{4}(3 \sin \omega t - \sin 3\omega t)\right] \quad (35)$$

The fundamental component is

$$\left[\frac{1}{\alpha}\left(\frac{w}{v_d}\right)\right]_{\text{fund.}} = A_\psi(1 + \tfrac{3}{4}A_\psi^2) \sin \omega t \quad (36)$$

Considering the ratio, output/input, as a transfer function, the characteristic forms a describing function for fundamental output for sinusoidal input given by

$$D(A_\psi) = 1 + \tfrac{3}{4}A_\psi^2 \quad (37)$$

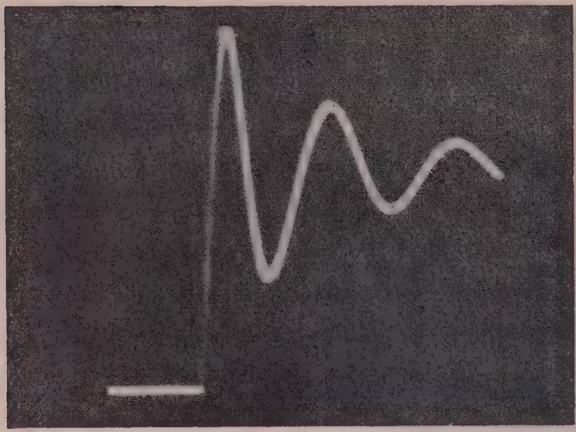
For convenience α is not included in the describing function but is included with the gain of the rest of the loop.

Using the fundamental transfer function the open-loop equation of the non-linear system becomes

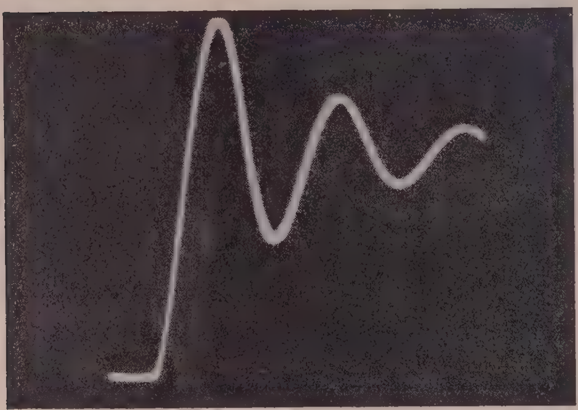
$$p^2 e_o = \frac{1}{T^2}\alpha G D(A_\psi)(1 + pT_{d1})e$$

$$\frac{e_o}{e} = -\frac{1}{T^2}\frac{\alpha G}{\omega^2}(1 + j\omega T_{d1})D(A_\psi) \quad (38)$$

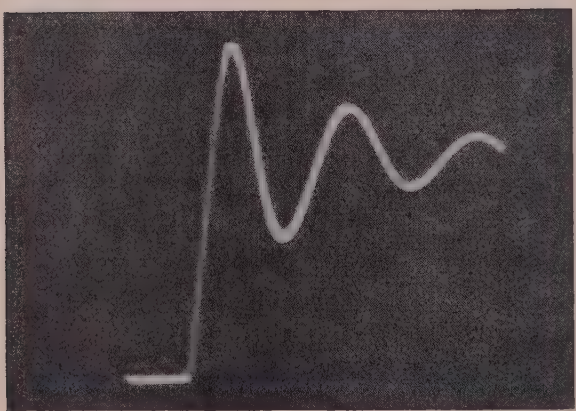
where it is assumed that $\xi T_{d1} = 0$



(a)



(b)



(c)

Fig. 16.—Experimental transient responses and using derivative-of-error stabilization.

$m = 5 \cdot 0.$
$D_e = 0 \cdot 20 D_{ec}.$

(a) Non-linear response, $E_i = 4 \cdot 3$.
(b) Corresponding linear response, $E_i = 4 \cdot 3$.
(c) Corresponding non-linear response using velocity feedback stabilization, $E_i = 4 \cdot 3$.

and $$p = j\omega$$

Eqn. (38) may be written in the form

$$\frac{e_o}{e} = D(A_\psi)\Phi(\omega)$$

The closed-loop response is given by

$$\frac{e_o}{e_i} = \frac{D(A_\psi)\Phi(\omega)}{1 + D(A_\psi)\Phi(\omega)}$$

$$\doteq \frac{\Phi(\omega)}{\dfrac{1}{D(A_\psi)} + \Phi(\omega)} \quad . \quad . \quad . \quad . \quad . \quad (39$$

The critical point for stability, the $(-1, 0)$ point in the linea system, is now a series of points lying on the locus $-1/D(A_\psi$ as A_ψ is varied. If the frequency locus $\Phi(\omega)$ encloses an portion of the amplitude locus $-1/D(A_\psi)$ for $0 < A_\psi < \infty$ the closed-loop system will be unstable.

These two curves are shown in Fig. 18, and it is seen that for a second-order system no instability can occur for any amplitud of input signal.

In actual systems the required form of the transfer functio $(1 + PT_{d1})$ of the phase-advance unit is not realizable and ther is an unwanted time-constant ξT_{d1}.

In many servo-systems the signals are discontinuous and a smoothing circuit must be used, the simplest form having a transfer function

$$\frac{1}{1 + pTs}$$

If these additional time-constants are included, the locus o $\Phi(\omega)$ may cut the real axis near the origin and the describin function at a point P is as shown in Fig. 18. P correspond to an unstable limit cycle of such a form that a displace ment to the left produces a stable system and a displacemen to the right produces an unstable system with ever-increasin amplitude of oscillation. However, saturation will produce a stable limit cycle.

In practice, saturation of the cubic non-linearity will occur a some amplitude. The corresponding describing function is shown in Fig. 18.

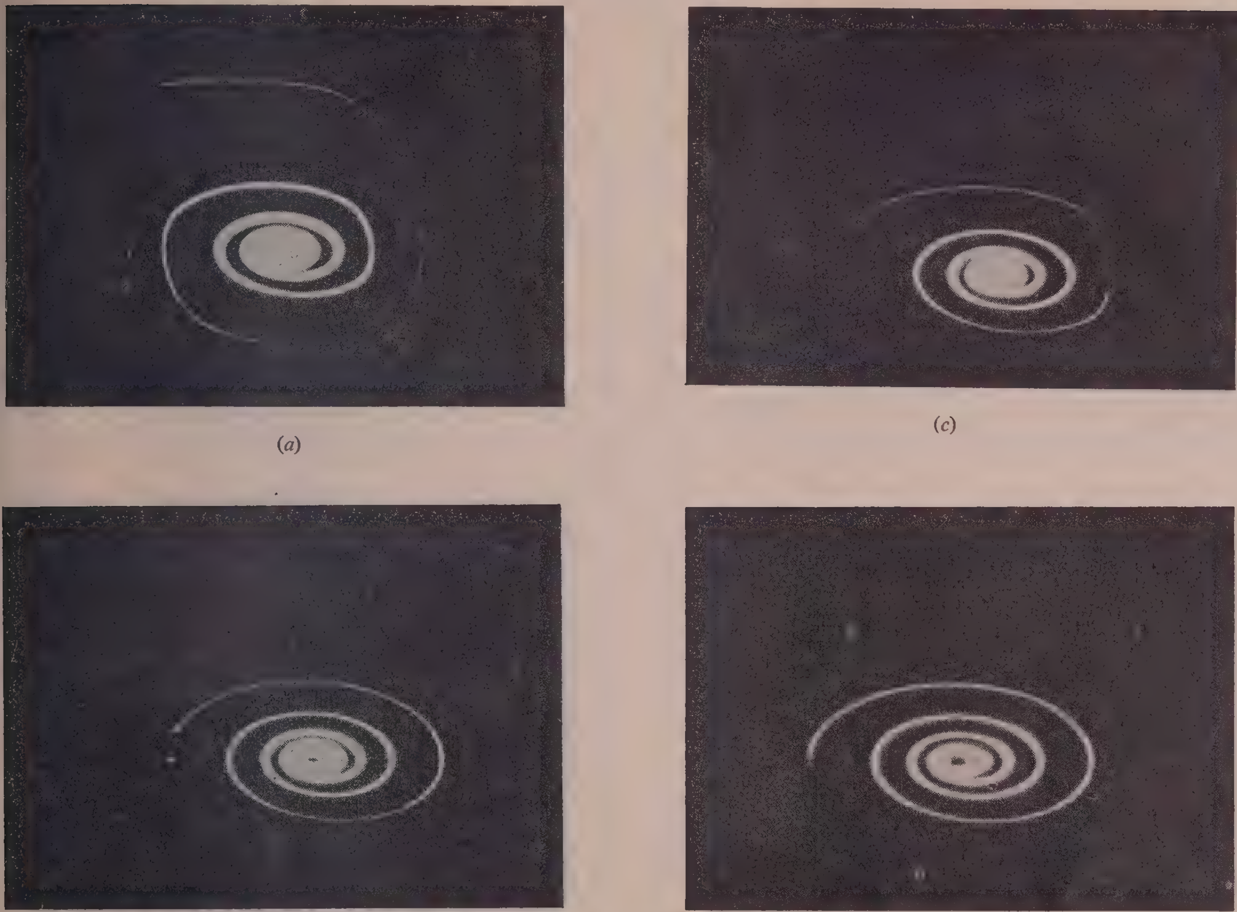
In this case the stability conditions are as shown by curves (b), (c) and (d) of Fig. 18. The amplitude locus is curve (b). The frequency locus for a small gain factor is represented by curve (d) whilst the frequency locus for a larger gain factor is represented by curve (c). The frequency locus represented by curve (d) will give rise to a stable closed loop system irrespective of amplitude, whilst the locus represented by curve (c) will give rise to two limit cycles, one divergent and one convergent.

(7.2) The Stability of the System Possessing the Segmented-Line Characteristic

The describing function for the segmented-line characteristic which is derived in Reference 1 (Appendix), is given by the expression

$$D(B) = k + \frac{2m_a}{\pi}\left\{\left(\frac{\pi}{2} - \phi\right) - \frac{v_c}{B}\sqrt{\left[1 - \left(\frac{v_c}{B}\right)^2\right]}\right\}$$

where $\phi = \arcsin(v_c/B)$ is the angle at which the non-linearity first occurs and B is the amplitude of the input signal to the non-linearity. The critical point of stability is now a series of points lying on the locus of $-1/D(B)$. This locus lies along the negative real axis between the points $(-1, 0)$ and $(-k/m, 0)$ and is shown in Fig. 18; the critical point for a linear system with an amplifier of gain k being $(-1, 0)$.



(a) (c)

(b) (d)

Fig. 17.—Experimental phase-plane.

$m = 5{\cdot}0.$
$D_e = 0{\cdot}20 D_{ec}.$

(a) Non-linear trajectory using derivative-of-error stabilization.
(b) Corresponding linear trajectory.

(c) Non-linear trajectory using velocity feedback stabilization.
(d) Corresponding linear trajectory.

Curve (c) in Fig. 19 represents the frequency locus of a second-rder system. No instability can occur for any amplitude of the nput signal. Curves (b) and (a) represent typical frequency loci f a fourth-order system for two different values of gain factor. he locus for small gain represented by curve (b) will give rise o a stable closed-loop system irrespective of amplitude; the ocus for a larger gain factor represented by curve (a) will give ise to an unstable closed-loop system.

) THE TRANSIENT BEHAVIOUR OF A LINEAR SYSTEM WITH AN AMPLIFIER OF GAIN M

The transient response of the system with the segmented-line haracteristic for critical damping in the non-linear regime is hown in Fig. 20(a), and the response of the linear system obtained hen the non-linear element is replaced by a linear amplifier of ain M is shown in Fig. 20(b).

For this damping the difference is most evident, especially for mall amplitudes of input signal. For large inputs there is little ifference between the responses.

(8.1) Stability Considerations

It is possible that a linear system of order higher than the second will be unstable at a gain M_1. For example, the frequency locus of a fourth-order linear system is shown in Fig. 18 to pass through the critical point $(-k/m_1, 0)$; the critical point for a linear system with an amplifier of gain k being $(-1, 0)$. If there is no saturation the amplitude locus of the corresponding segmented-line characteristic lies between the points $(-1, 0)$ and $(-k/m_1, 0)$. Because of saturation, the non-linear system will be stable at all amplitudes of the input signal; moreover, the non-linearity can be increased considerably before instability occurs.

(8.2) Experimental Results

The second-order linear system was changed to a fourth-order system by the addition of two extra time-constants ξT_{d1} and T_s. The first was introduced by increasing the unwanted time-constant of the phase-advance unit and the second was introduced by the addition of a smoothing circuit.

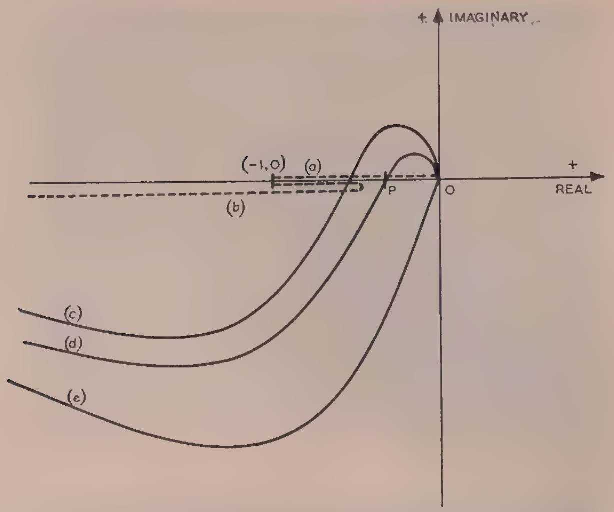


Fig. 18.—Nyquist diagram for a system with a cubic characteristic.

(*a*) Negative reciprocal of the describing function for a cubic characteristic.
(*b*) Negative reciprocal of the describing function for a cubic characteristic when saturation is present.
(*c*) Frequency locus of a fourth-order system.
(*d*) Frequency locus of a fourth-order system, with smaller gain factor than that of (*c*).
(*e*) Frequency locus of a second-order system.

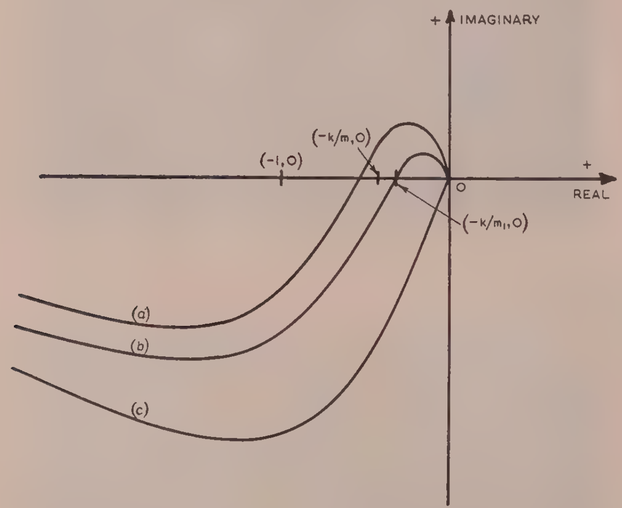


Fig. 19.—Nyquist diagram for a system with a segmented-line.

(*a*) Frequency locus of a fourth-order system.
(*b*) Frequency locus of a fourth-order system, with a smaller gain factor than that of (*a*).
(*c*) Frequency locus of a second-order system.

The open-loop equation of this fourth-order linear system is

$$p^2T^2e_o = MG\left[\left(\frac{1+pT_{d1}}{1+p\xi T_{d1}}\right)\Big/(1+pT_s)\right]e$$

and the closed-loop transfer function is

$$\frac{e_o}{e} = \frac{MG(1+pT_{d1})}{a_0p^4 + a_1p^3 + a_2p^2 + a_3p + a_4} \quad . \quad . \quad (40)$$

where $$a_0 = \frac{T_s\xi T_{d1}}{\omega_n^2}$$

$$a_1 = \frac{T_s + \xi T_{d1}}{\omega_n^2}$$

$$a_2 = \frac{1}{\omega_n^2}$$

$$a^3 = MGT_{d1}$$

$$a_4 = MG$$

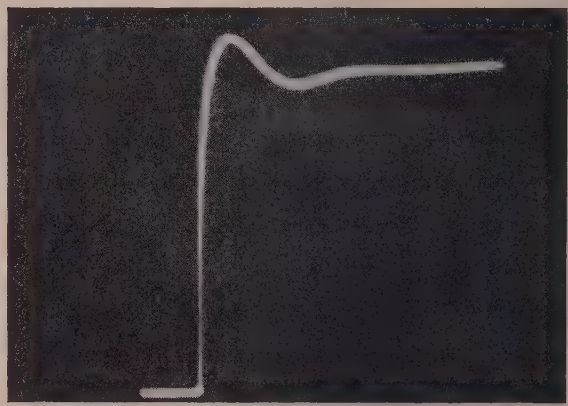
Using the following constants,

$$T_s = \xi T_{d1} = \tfrac{1}{10}T_{d1}$$

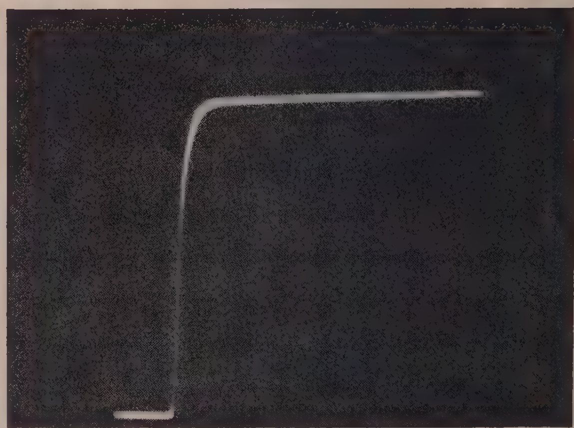
$$\omega_n T_{d1} = 2\cdot 4$$

$$G = 0\cdot 44$$

it is found from the Hurgitz–Routh stability criterion that for stability $0 < m < 6\cdot 4$. It was found experimentally that the system became unstable at a value of $m = 6\cdot 2$. The equivalent



(*a*)



(*b*)

Fig. 20.—Experimental transient responses of a system using derivative-of-error stabilization.

$m = 5\cdot 0$.
D_e is critical in the non-linear regime.
$E_i = 6\cdot 8$.

(*a*) Non-linear response.
(*b*) Response of the linear system with an amplifier of gain $M = 5$ replacing the amplifier of gain $k = 0\cdot 94$.

non-linear system was stable for all amplitudes of the input signal. Furthermore, the non-linear system remained stable for all amplitudes of input signal up to a value of $m = 9 \cdot 3$.

(9) THE NECESSITY FOR EXPERIMENTAL TESTING WITH RANDOM INPUTS

For the design of linear servo mechanisms there exist formal mathematical relationships[15,16,17] between the response to one form of input and the response to any other input; the three commonest inputs considered are sinusoidal, transient and random. Although for a specific case these relationships may be difficult to evaluate, long usage of frequency-response and transient-response testing has given the engineer a qualitative idea of these relationships. In particular, it is known that the larger the bandwidth of a steady-state frequency-response characteristic, the shorter in time will be the transient response to step or impulse displacement, the resonant peak in the frequency characteristic giving some indication of damping of the transient. In addition, the frequency-response characteristic obtained by the use of a single sinusoid shows how the system will behave when the input is compounded of many frequency components. The bandwidth also implies that the system will respond satisfactorily to any arbitrary input signal provided that the frequency spectrum of the input does not exceed the bandwidth of the system.

Unfortunately, few mathematical relationships[18] exist for the non-linear systems, and furthermore insufficient experimental work[20] has been carried out on these systems to enable an engineer to predict qualitatively, knowing the response to only one input, the response to any other input. Since the behaviour of a particular non-linear system is not representative of all non-linear systems, the design, construction and testing of a system general enough to include all types of non-linearities is a formidable problem. However, limiting the experimental system to a particular class should reveal whether, for this particular class, a qualitative relationship exists between these three forms of testing, and furthermore whether this relationship agrees favourably with that which exists for the linear system.

The system described in Section 3 lends itself to such testing, since it is representative of a particular class of non-linear systems possessing one frequency-independent single-valued non-linearity and being integrator smoothed. Work on the steady-state frequency response of this system has shown that the bandwidth increases and the magnitude of the resonant peak decreases with increasing signal amplitude and/or increasing non-linearity. In the first part of the paper it has been shown that for step displacements the percentage overshoot and rise time decrease for similar conditions. The aim of the work is to obtain the response of the system to random signals, to determine whether a qualitative relationship exists for these three forms of testing and finally to determine whether this relationship, if it exists, compares favourably with that of the linear system.

(10) THE EFFECT OF A NON-LINEARITY ON EXPERIMENTAL TECHNIQUES

When the input θ_i to the system is composed of signal S_i and noise N_{in}, the response S_o to S_i in the absence of N_{in} is modified by the presence of N_{in} and vice versa. Consequently the signal/noise ratio at the output has no significant meaning. The testing of the system is complicated, therefore, not only by the fact that the non-linearity is amplitude dependent but also by the cross-modulation which takes place. In this work the signal/noise ratio at the input is well defined since experimentally signal and noise are independent sources and are measured separately. The input to the system is defined by assuming the signal and the signal/noise ratio to be independent parameters. Without further investigation it is not possible to relate the amplitude of the output spectrum about a frequency f to the amplitude of the input spectrum at the same frequency because of the cross-modulation. An inspection of the error spectrum and the output spectrum gives an indication of the amount of cross-modulation, and it will be seen experimentally that for this particular system the amount is negligible.

(10.1) The Noise Spectra

The noise output of a 931–A photo-multiplier,[19] a simple arrangement of which is described in Appendix,[14] was used as a source of white noise, the spectrum remaining flat beyond the cut-off frequency of the system. Noise with four flat spectra of limited bandwidth was obtained by passing white noise through filters, the respective half-power points* occurring at 500, 1 500, 3 000 and 5 000 c/s. Narrow-band noise was obtained by passing white noise through a selective amplifier with a pass band of 150 c/s centred at 1 025 c/s. Noise with a clipped spectrum was obtained by clipping white noise at $\pm 1 \cdot 5$ volts.

(10.2) Spectrum Analysis

The voltage spectrum of the noise was determined by passing the noise through a tunable selective amplifier with a constant pass band and followed by measuring and smoothing circuits. Selective filters with a pass band of 30 c/s were used. This necessitated a waiting time of less than one minute to obtain an accurate reading, and the accuracy of the readings was about 5%. Filters with very narrow pass bands were found unsuitable owing to the long waiting time required to obtain an accurate reading.

(11) EXPERIMENTAL RESULTS

The contribution of the cross-modulation to the output is shown in Figs. 21 and 22. The relative magnitude of the first 23 harmonics in the linear and non-linear response to a 50/50 square wave are compared in Fig. 21. The responses to a noise signal with the half-power point at 500 c/s are shown in Fig. 22. In both Figures the linear and non-linear responses are very nearly the same, which indicates negligible cross-modulation.

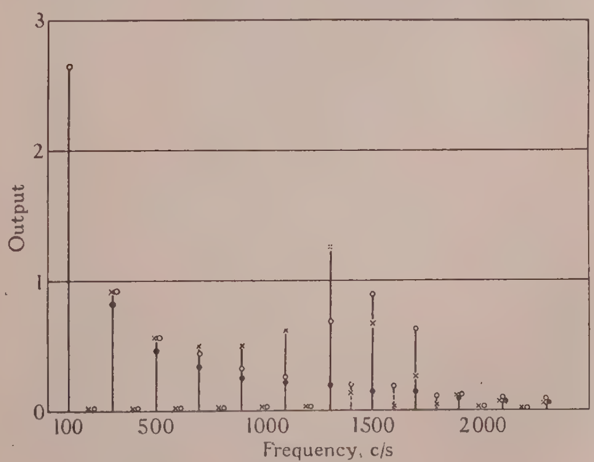


Fig. 21.—The response to a 100 c/s square wave showing the relative amplitudes of the harmonics.

Input spectrum.
Linear response.
Non-linear response.

* The half-power point is defined as that point at which the noise voltage/unit-bandwidth is one-half its initial value.

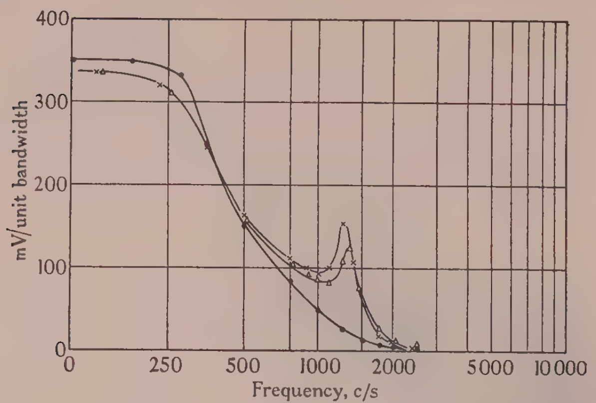


Fig. 22.—The response to a random signal with a low-frequency power spectrum.

$T_{d1} = 0 \cdot 16 T_{d1c}$

● Power spectrum of the input signal.
× Power spectrum of the output signal, linear response.
▽ Power spectrum of the output signal, non-linear response.

$M = 7 \cdot 7$, $N_{is} = 2 \cdot 0$

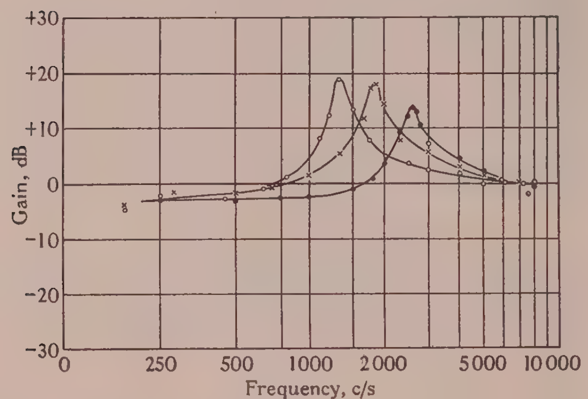


Fig. 23.—Error response to white noise showing the effect of varying the signal power.

$T_{d1} = 0 \cdot 16 T_{d1c}$

○ Linear response.
× $M = 4 \cdot 1$, $N_{is} = 0 \cdot 56$
● $M = 4 \cdot 1$, $N_{is} = 2 \cdot 2$

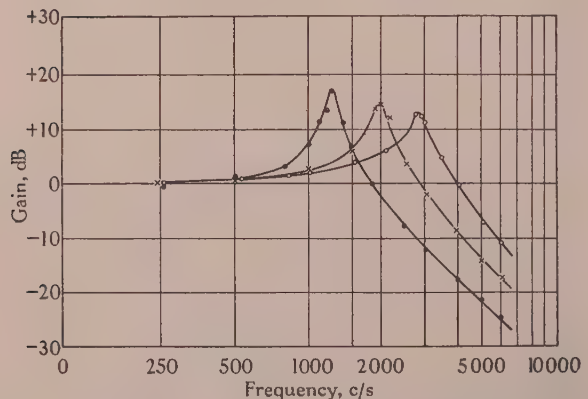


Fig. 24.—Output response to white noise showing the effect of varying the non-linearity.

$T_{d1} = 0 \cdot 16 T_{d1c}$

● Linear response.
× $M = 4 \cdot 1$, $N_{is} = 1 \cdot 75$
○ $M = 7 \cdot 7$, $N_{is} = 1 \cdot 75$

The response to white noise is shown in Figs. 23 and 24. The error spectrum is shown in Fig. 23 and the output spectrum in Fig. 24. The effect of increasing the non-linearity and/or the noise signal is to increase the bandwidth of the response and to reduce the resonant peak.

The response to a sinusoidal signal in the presence of white noise is shown in Figs. 25 and 26. Fig. 25 shows that the effect

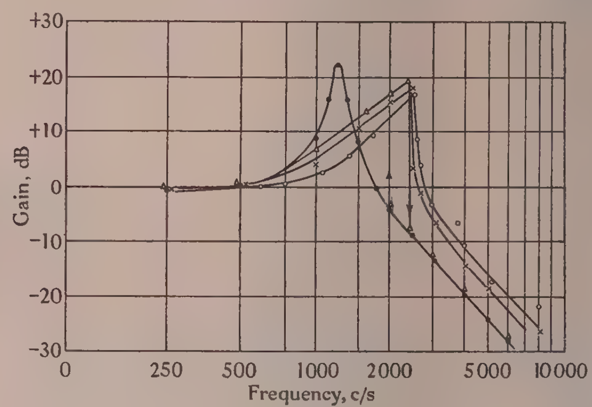


Fig. 25.—Output response to a sinusoidal signal in the presence of white noise.

$T_{d1} = 0 \cdot 08 T_{d1c}$

● Linear response.
△ $M = 4 \cdot 1$, $S_i = 1 \cdot 0$, $N_{in} = 0 \cdot 013$
× $M = 4 \cdot 1$, $S_i = 1 \cdot 0$, $N_{in} = 1 \cdot 55$
○ $M = 4 \cdot 1$, $S_i = 1 \cdot 0$, $N_{in} = 2 \cdot 00$

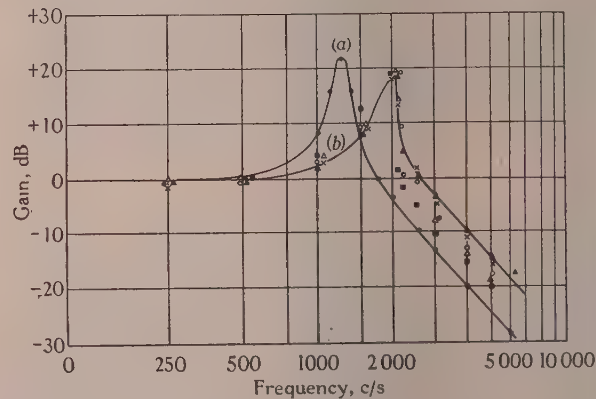


Fig. 26.—Output response to a sinusoidal signal in the presence of random disturbances.

$T_{d1} = 0 \cdot 08 T_{d1c}$, $M = 4 \cdot 1$, $S_i = 0 \cdot 87$, $N_{in} = 1 \cdot 75$

(a) Linear response.
(b) Non-linear response.

▲ White noise.
× Noise with a clipped spectrum.
■ Narrow-band noise centred at 1 025 c/s.
○ Noise with the half-power point at 3 000 c/s.
△ Noise with the half-power point at 1 500 c/s.

of increasing the noise while keeping the signal constant is to quench the jump phenomena in the frequency characteristic, to reduce the resonant peak and to increase the bandwidth. When the disturbance is small compared with the signal, the response is independent of the bandwidth of the disturbance. When the disturbance is comparable with the signal, limiting the bandwidth of the disturbance while keeping the power of the disturbance constant affects the sinusoidal response, increasing the attenuation with a decrease in the bandwidth of the disturbance. This is shown in Fig. 26.

The effect of white noise on the discontinuous jumps in the frequency characteristic is shown in Fig. 27. The noise affects appreciably only the jumps that occur for decreasing signal frequency, the effect becoming more pronounced as the non-linearity is increased. This is shown by curves (*a*) and (*b*) of

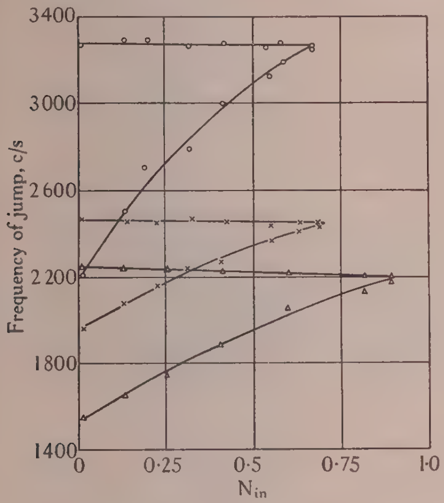


Fig. 27.—Variation of jump frequencies with noise power.

(*a*) —○— $M = 7 \cdot 7$, $S_i = 1 \cdot 21$
(*b*) —×— $M = 4 \cdot 1$, $S_i = 1 \cdot 20$
(*c*) —△— $M = 4 \cdot 1$, $S_i = 1 \cdot 00$

Fig. 27. It has also been found that for any degree of non-linearity there is a particular signal amplitude at which the jump phenomena may be retained in the presence of a maximum amount of noise.

The response to a random signal with a limited bandwidth, the half-power point being at 5000 c/s, is shown in Fig. 28.

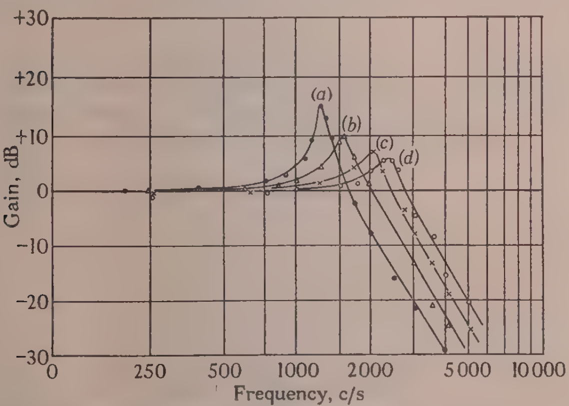


Fig. 28.—Output response to a low-frequency-bandwidth random signal in the presence of white noise.

$T_{d1} = 0 \cdot 16 T_{d1c}$, $M = 4 \cdot 1$
—●— Linear response.
—△— $N_{is} = 0 \cdot 67$, $N_{in} = 0 \cdot 00$
—×— $N_{is} = 0 \cdot 67$, $N_{in} = 0 \cdot 33$
—○— $N_{is} = 0 \cdot 67$, $N_{in} = 0 \cdot 67$

Curves (*a*) and (*b*) are the responses to the signal alone, and curves (*c*) and (*d*) are the responses to the signal in the presence of white noise. The effect of increasing the signal and/or the noise is to increase the bandwidth and to reduce the resonant peak of the non-linear response. A similar effect is observed when the non-linearity is increased, the input being kept constant.

(12) CONCLUSIONS

The transient response of a system stabilized by output-velocity feedback and incorporating a hard-spring amplitude characteristic has been analysed graphically for two forms of the characteristic. The salient features of the non-linear response have been discussed and compared with the response of the equivalent linear system. It has been shown that under certain conditions the response of the former is better than that of the latter. The effect of phase-advance stabilization has been derived, and the response of a system incorporating this stabilization has been compared with that of a system stabilized by velocity feedback. It has been shown that the response time of a system stabilized by phase advance is less than that of a system stabilized by velocity feedback, and furthermore that the first overshoot of the former is greater than that of the latter. The effect of the non-linear amplitude characteristic on the overall stability of the system has been discussed, and a comparison has been made of the stability of this system with that of a system incorporating a high-gain amplifier. It has been shown that under certain conditions the former is more stable. Experimental verification is included.

The effects of non-linearity on the experimental techniques necessary for testing non-linear control systems subjected to random inputs have been discussed, and experimental results obtained. It has been shown that the contribution to the output by cross-modulation products is negligible, and consequently the output spectrum can be related to the input spectrum to obtain a "gain-frequency characteristic." Thus the response to white-noise signals has been determined. It has been shown that the effect of increasing the signal and/or non-linearity is to increase the bandwidth and to decrease the resonant peak of the response. The effect of noise on the response to sinusoidal signals is to quench the jump phenomena if it is present, to reduce the resonant peak and to increase the bandwidth. It has been found that for this non-linearity there is a particular amplitude of sinusoidal signal at which a maximum amount of noise is required to quench the jump phenomena. This suggests that by pre-amplifying the input by a controlled amount a similar system properly designed might be used as a low-pass filter when the signal is relatively sinusoidal. Finally, the response of the system to a random signal in the presence of white noise has been found to be of the same character as the response to sinusoidal and white-noise inputs.

One of the chief aims of this work is to determine whether a qualitative relationship exists between frequency response, transient response and random signal testing. A comparison of the results in Section 11 with those in Section 5 and Reference 1 reveals that the increased bandwidth and decreased resonant peak of the sinusoidal response are related to an increased bandwidth and decreased resonant peak of the random signal response and to a quicker, less-oscillatory transient response. There is therefore a definite qualitative relationship between these three forms of testing, and this qualitative relationship agrees favourably with that which exists for the linear system.

(13) REFERENCES

(1) West, J. C., and Nikiforuk, P. N.: "The Behaviour of a Remote-Position-Control Servo Mechanism with Hard-Spring Non-Linear Characteristic," *Proceedings I.E.E.*, Paper No. 1621 M, March, 1954 (**101**, Part II, p. 481).

(2) McDonald, D. "Multiple Mode Operation of Servo-mechanisms," *Review of Scientific Instruments*, 1952, **23**, p. 22.

(3) West, J. C., and Dalton, I. R. "The Step-Function Response of an R.P.C. Servo Mechanism possessing Torque Limitation," *Proceedings I.E.E.*, Paper No. 1576 M, November, 1953 (**101,** Part II, p. 166).

(4) West, J. C., Douce, J. L., and Naylor, R. "The Effects of the Addition of some Non-linear Elements on the Transient Performance of a Simple R.P.C. System possessing Amplifier Saturation," *ibid.*, Paper No. 1549 M, August, 1953 (**101**, Part II, p. 156.

(5) West, J. C., and Douce, J. L.: "The Frequency Response of a Certain Class of Non-linear Feedback Systems," *British Journal of Applied Physics*, 1954, **5**, p. 204.

(6) Lienard, A.: "Etude des Oscillations Entretenues," 1928, *Revue Générale d'Électricité*, 1928, **23**, p. 901.

(7) West, J. C.: "Textbook of Servomechanisms" (English Universities Press Ltd., 1953), Chapter 21.

(8) Williams, F. C., and Ritson, F. J. U.: "Electronic Servo Simulators," *Journal I.E.E.*, 1947, **94**, Part IIA, p. 112.

(9) Ritson, F. J. U., and Hammond, P. H.: "The Design and Testing of an Electronic Simulator for a Hydraulic Remote-Position-Control Servo Mechanism," *Proceedings I.E.E.*, Paper No. 1296 M, March, 1952 (**99**, Part II, p. 533).

(10) Andronow, A. A., and Chaikin, C. E.: "Theory of Oscillations" (University Press, Princeton, 1949). Chapter 1.

(11) Michel, J. G. L., and Porter, A.: "The Effect of Friction on the Behaviour of Servo Mechanisms at Creep Speeds," *Proceedings I.E.E.*, Paper No. 1033 M, August, 1950 (**98**, Part II, p. 297).

(12) Wilkie, D. R.: "The Relation between Force and Velocity in Human Muscle," *Journal of Physiology*, 1950, **110**, Nos. 3 and 4, p. 249.

(13) Kochenburger, R. I.: "A Frequency Response Method for Analysing and Synthesizing Contactor Servo-mechanisms," *Transactions of the American I.E.E.*, 1950, **69**, Part 1, p. 270.

(14) Blumenthal, J. S., and Beck, F. J.: "Transient Analysis of Non-linearized Single Lag Servomechanisms," *Proceedings of the U.S. National Congress of Applied Mechanics*, 1952.

(15) Pelegrin, M. J.: "Application of the Statistical Technique to the Servo-Mechanisms Field" (Automatic and Manual Control, Butterworths Scientific Publications, 1952), p. 123.

(16) Jones, S.: "The Determination of the Best Form of Response for a Servo when an Extraneous Random Disturbance is Present with the Error Signal" (Automatic and Manual Control, Butterworths Scientific Publications, 1952), p. 139.

(17) James, H. M., Nichols, N. B., and Phillips, R. S.: "Theory of Servo-Mechanisms" (McGraw-Hill, 1947), Chapter 8.

(18) Booton, R. C.: "Non-Linear Control Systems with Statistical Inputs," Report No. 61. Dynamic Analysis and Control Laboratory, Massachusetts Institute of Technology.

(19) Sard, R. O.: "Calculated Frequency Spectrum of the Shot Noise from a Photomultiplier Tube," *Journal of Applied Physics*, 1946, **17**, p. 768.

(20) Hammond, P. H., and Smith, R. A.: "An Experimental Study of the Response of On-Off Controlled Servo-Mechanisms to Periodic and Randomly Varying Input Signals," T.R.E. Technical Note No. 178, December, 1952.

(14) APPENDICES

(14.1) The Derivation of the Normalized Equation of Motion of a Simulator possessing a Cubic Non-Linear Characteristic

Let θ_i and θ_o be the actual instantaneous position of input and output.

Let $$\theta_i = v_d e_i$$

and $$\theta_o = v_d e_o$$

Then the normalized processed error signal is given by

$$e = G(e_i - e_o)$$

and the normalized control signal by

$$e_c = G(e_i - e_o) - T_d \dot{e}_o$$

The output from the non-linear element is of the form

$$w/v_d = \alpha(\psi + \psi^3)$$

If it is assumed that the friction is zero, and that the system is of second order, the equation of motion is given by

$$T^2 \ddot{e}_o = \alpha[(e - T_d \dot{e}_o) + (e - T_d \dot{e}_o)^3] \quad . \quad . \quad (41)$$

For convenience in the following analysis it is assumed that the input is held at zero level and that the output is released from rest with an initial error $-|\theta_{o1}|$. For this simulator, but not in general, this response is the same as the step-function response with the input moving from $\theta_i = -|\theta_{o1}|$. In the phase-plane analysis it is preferable to use the normalized form of the equation of motion. Let the input be held at zero and let

$$s = \frac{1}{T} t$$

$$\omega_n = \frac{1}{T}$$

$$Y = \frac{de_o}{ds}$$

$$\dot{e}_o = \omega_n Y$$

$$-X = e_i - e_o \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (42)$$

Substituting eqns. (42) in eqn. (41) and simplifying the result, one obtains the normalized equation of motion

$$Y\frac{dY}{dX} = -\alpha[(GX + T_d \omega_n Y) + (GX + T_d \omega_n Y)^3] \quad . \quad (43)$$

(14.2) Determination of the Normalized Equations of Motion of a Simulator possessing a Segmented-Line Characteristic

The three conditions under which the simulator possessing this particular characteristic operates are given in Section 4.

Let $$\theta_i = V_c e_i$$

and $$\theta_o = V_c e_o$$

The normalized processed error is given by

$$e = G(e_i - e_o)$$

and the normalized control signal is given by

$$e_c = G(e_i - e_o) - T_d \dot{e}_o$$

If it is assumed that the friction is zero and that the system is of second order, the equations of motion corresponding to each mode of operation are

$$T^2 \ddot{e}_o = k[G(e_i - e_o) - T_d \dot{e}_o] \quad . \quad . \quad . \quad . \quad (44)$$

for condition 1

$$T^2 \ddot{e}_o = m[G(e_i - e_o) - T_d \dot{e}_o] - (m - k) \quad . \quad . \quad (45)$$

for condition 2

and $$T^2 \ddot{e}_o = m[G(e_i - e_o) - T_d \dot{e}_o] + (m - k) \quad . \quad . \quad (46)$$

for condition 3

If for convenience it is assumed that θ_i is held at zero, the normalized equations of motion may be obtained by substituting eqns. (42) in the above three equations of motion.

The resulting normalized equations of motion are

$$Y\frac{dY}{dX} + \omega_n T_d kY + kGX = 0 \quad . \quad . \quad . \quad (47)$$

for condition 1

$$Y\frac{dY}{dX} + \omega_n T_d k(m/k)Y + mGX = -\delta \quad . \quad . \quad (48)$$

for condition 2

$$Y\frac{dY}{dX} + \omega_n T_d k(m/k)Y + mGX = \delta \quad . \quad . \quad (49)$$

for condition 3

where $$\delta = (m - k) = m_a \quad . \quad . \quad . \quad . \quad . \quad . \quad (50)$$

(14.3) Determination of the Straight-Line Trajectories in the Non-Linear Regime

The system described in Section 3 is critically damped in the linear regime and overdamped in the non-linear regime. There are then two straight-line trajectories corresponding to each condition of operation in the non-linear regime.

For the condition $V_p > V_c$ the normalized equation of motion is

$$Y\frac{dY}{dX} + \omega_n T_d k(m/k)Y + mGX = -\delta$$

The equation of the trajectory will be the same as the equation of the locus.

In the above equation let

$$S = \frac{dY}{dX} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (51)$$

and $$Y = S\left[X + \frac{1}{G}\left(1 - \frac{k}{m}\right)\right] \quad . \quad . \quad . \quad . \quad (52)$$

which is the equation of a straight line passing through the point

$$X = -\frac{\delta}{mG}, \; Y = 0$$

The required slope will then be given by

$$S^2 + \omega_n T_d k(m/k)S + mG = 0$$

This has the solution

$$S_1 = \frac{-\omega_n T_d k(m/k) - \sqrt{\{[\omega_n T_d k(m/k)]^2 - 4mG\}}}{2} \quad . \quad (53)$$

and $$S_2 = \frac{-\omega_n T_d k(m/k) + \sqrt{\{[\omega_n T_d k(m/k)]^2 - 4mG\}}}{2} \quad . \quad (54)$$

For the condition $V_p < -V_c$ the slopes S_1' and S_2' of the straight-line trajectories can be determined in a similar manner. The solutions are

$$S_1' = \frac{-\omega_n T_d k(m/k) - \sqrt{\{[\omega_n T_d k(m/k)]^2 - 4mG\}}}{2} = S_1$$

and

$$S_2' = \frac{-\omega_n T_d k(m/k) + \sqrt{\{[\omega_n T_d k(m/k)]^2 - 4mG\}}}{2} = S_2$$

If the damping in the linear regime is reduced to a value

$$\omega_n T_d k = 2\sqrt{kG}\sqrt{(k/m)}$$

the system is critically damped in the non-linear regime, there is but one straight-line trajectory corresponding to each condition of operation in the non-linear mode, and

$$S_1 = S_2 = -\frac{\omega_n T_d k(m/k)}{2} \quad . \quad . \quad . \quad . \quad (55)$$

For a value of damping less than this, the system is under-damped in both regimes.

(14.4) Determination of the Intersections of the Loci of Constant Slope S with a Boundary

The locus of constant slope S'' for the condition $|V_p| < V_c$ is given by

$$Y = \frac{-kGX}{S'' + \omega_n T_d k} \quad . \quad . \quad . \quad . \quad . \quad (56)$$

If the point at which this locus intersects the boundary

$$Y = -\frac{1}{T_d \omega_n}(GX + 1)$$

is X_1, Y_1, then the solution is

$$X_1 = -\frac{S'' + \omega_n T_d k}{S''G}, \; Y_1 = \frac{k}{S''}$$

The locus of constant slope S'' for the condition $V_p > V_c$ is given by

$$Y = -\frac{mGX + \delta}{S'' + (\omega_n T_d k)(m/k)} \quad . \quad . \quad . \quad (57)$$

If the point at which this locus intersects the boundary

$$Y = -\frac{1}{T_d \omega_n}(GX + 1)$$

is X_2, Y_2, then the solution is

$$X_2 = -\frac{S'' + \omega_n T_d k}{S''G} = X_1$$

$$Y_2 = k/S'' = Y_1.$$

Hence the locus of constant slope in the non-linear regime intersects a boundary at the same point as the locus of the same slope in the linear regime.

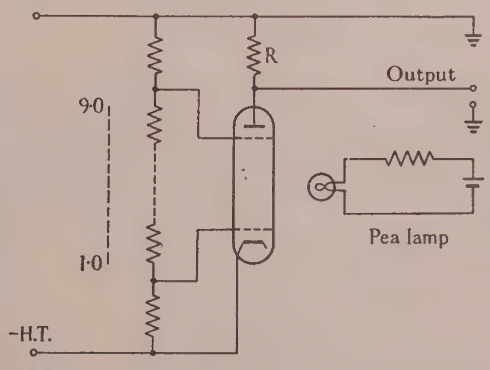


Fig. 29.—Arrangement of the photo-multiplier.

(14.5) The Noise Source

The arrangement used is shown in Fig. 29, the photo-multiplier tube and the pea lamp being enclosed in a metallic cylinder. It was found that, using a well-stabilized power supply and a good accumulator for the light source, the noise output remained constant to within a few per cent for periods of half an hour and longer.

DISCUSSION BEFORE THE MEASUREMENTS SECTION, 15TH MARCH, 1955

Dr. J. H. Westcott: The authors have based their papers on the description-function method, and I believe that Daniel first suggested its use. I should like to see a disclaimer, stating that the paper is based on the description-function method and resemblance to non-linear analysis is so limited. We are steeped in linear theory, and this is a device for enabling us to do something in cases where otherwise we could do nothing, and so it is useful.

I am interested in the phase-margin criterion of 21°. Once we were told to keep outside the limit of 30°, and then we were told that we could do better than this. Now the authors categorically advocate keeping out of the limit once more.

With regard to Fig. 7 of the paper by Drs. West and Douce, the authors give some calculated values showing how good their experimental values are. I do not know how they were calculated, but suspect it was on the basis of the description-function theory. The experimental points are then accurate and the calculated points are approximate, but I am not sure of the accuracy of the experimental points because I am not sure how far we can trust the simulator. This raises the question of how we can successfully record computer work.

Mr. P. E. W. Grensted: A system very similar to that considered in the paper by Drs. West and Nikiforuk is a velocity-lag servo mechanism with a non-linear gain characteristic which is an nth power law. The equation for the output x in response to a step input is

$$\frac{d^2x}{dt^2} + 2c\frac{dx}{dt} + \{x\}^n = 0, \; n \geqslant 0$$

where $\{x\}^n = x^n$ if $x > 0$, and $\{x\}^n = -(-x)^n$ if $x < 0$.

The solution is oscillatory for large amplitudes if n is greater than unity (hard spring) and oscillatory for small amplitudes if n is less than unity (soft spring). It may be shown by an extension of the describing-function method* that the amplitude a of these oscillations is given very closely by

$$a = A\varepsilon^{-4ct/(3+n)}$$

where A is a constant. Thus the damping of the transient is given by

$$\mu = \frac{4}{3+n}c \quad \text{which is constant.}$$

The damping is increased from the value $\mu = c$ for a linear system ($n = 1$) if n is less than unity and decreased if n is greater than unity. For this system it appears that a soft-spring gain characteristic would be preferable.

With reference to the paper by Drs. West and Douce, does the theoretical analysis indicate whether the sub-harmonics are stable?

Prof. A. Tustin: Over quite a number of years Dr. West and his colleagues have been tackling some very awkward problems in non-linearities and presenting the results in an excellent sequence of papers. The work is considerably restricted by what is at present mathematically practicable. How far do the authors think that analysis on these lines may ultimately be extended to cater for the more complex systems used in practice? Phase-plane methods are limited to second-order systems, and the general nature of the conclusions given in the paper by Drs. West and Nikiforuk were self-evident from the start. It is unfortunate that similar results cannot be obtained for higher-order systems when we deal with step-functions. The greater part of the series of papers have dealt with functions of frequency, using the so-called "describing-function" approach. This method is very powerful, and it has given a great wealth of useful results. I believe that it can give most of what the engineer requires. The authors are not being quite fair in stating that the describing-function technique neglects harmonics. It does not exactly neglect harmonics; it simply assumes that if a periodic oscillation of a system is to occur, and if we consider the fundamental of that oscillation, it must, in the presence of the other harmonics, regenerate itself precisely. The same statement applies to each separate harmonic. To pick out the fundamental and draw its Nyquist diagram is not to neglect harmonics but to make the supposition that the presence of the harmonics, which are undoubtedly present, does not significantly modify the relationship between components of fundamental frequency.

I have never been able to understand why solutions for the response to step-functions—apart from their theoretical interest—should engage the time of so many people. Few servo mechanisms have to follow step-function inputs except under running-in conditions, and then the error is so large that in practice we could not obtain a cubic law for the response of torque to error. Its character is inevitably just the reverse. In such cases, I should like to know the practical use of a detailed inquiry into step-function responses for servo-mechanism design.

I have tried to follow the reference to the treatment of these problems in terms of bandwidth. The general results, which are described as having been arrived at through consideration of bandwidth, are evident from elementary physical principles. The higher frequency at larger amplitudes occurs just because the control is effectively stiffer. The peculiar shape of the response to a step-function is entirely due to the non-linearity, and bandwidth is a concept essentially associated with linearity and harmonic analysis, but I may have missed the point.

Having acquired so much experience in applying phase-plane and other methods to these non-linear problems, could the authors indicate the direction in which they feel further research in the field would be fruitful? Have they now reached the end of the problems which can be solved by mathematical methods? If so, what do they suggest we should do about it, since designers are concerned with non-linearities in high-order systems?

Mr. D. V. Blake: I would like to deal with the use of the term "better response" in the paper by Drs. West and Nikiforuk. The fast response is not necessarily the best. At the N.P.L. we have tackled a number of problems on the simulator with various customers, and each has his own idea of what is best. For instance, high speeds and large changes of speed are very bad in things like lifts and devices carrying people.

The paper would have been much more valuable if the authors had given an adequate comparison between their non-linear systems and a corresponding linear one. Most of the comparisons are given with reference to a linear system of lower gain, i.e. one corresponding to the first part of the non-linear characteristic. It would be helpful if they could give more information on the comparison with respect to a linear system with optimum gain—probably an intermediate value between the two used in the non-linear case. In this case, is the additional complication of providing a cubic characteristic justified?

At present, there seems to be a lack of adequate means of checking simulator solutions. In a linear system we can usually assume that it is reasonable and the response is approximately what would be expected. It can be checked at spot points by varying parameters and finding out the point at which instability occurs. This can be done fairly easily in a linear system, but in a non-linear one it is very difficult.

Mr. J. J. Thompson: The organization with which I am asso-

* GRENSTED, P. E. W.: "The Frequency Response Analysis of Non-linear Systems," *Proceedings I.E.E.*, Monograph No. 126 M, April, 1955 (102 C, p. 244).

ciated has used non-linear damping of the segmented line pattern in hydraulically-operated steam-turbine governors for the past 25 years, for the express purpose mentioned in the Introduction to the paper by Drs. West and Nikiforuk, namely that of decreasing the response time whilst maintaining stability. This feature is most useful in difficult governing problems, such as governing the rate of export or import of electrical power to or from a large network, from an industrial turbo-alternator which also feeds a local fluctuating load. In an a.c. system this form of governing involves, of course, not the speed control but the position control of the alternator rotor. The system equation thus becomes of higher order than usual, and in practice it is sometimes difficult to achieve stability without serious sacrifice of response time. The use of non-linear damping restores the response time to a normal value.

I must protest against the continued use of the term "hard spring." Because it was apparently used by a French author in 1928 and has since been used in the United States is no reason for using it here. If electrical engineers must use mechanical terms they should study accuracy. Hardness is by definition the ability to resist indentation. The hardness of a spring therefore has no effect on the characteristics of a spring as such, but rather on its fatigue limit. At the same time, why is the spring used as an example when other devices can exhibit an equally flexible relationship between input and output whilst having input and output expressed in the same units, which is not true of the spring? Fig. 5, showing a so-called "hard-spring" characteristic, corresponds to a spring of decreasing rate, which is most confusing. It would have been appropriate to refer to it as a "soft spring."

THE AUTHORS' REPLY TO THE ABOVE DISCUSSION

Drs. J. C. West, J. L. Douce and **P. N. Nikiforuk** (*in reply*): We are pleased to learn that, contrary to general knowledge, the introduction of describing-function techniques occurred in this country and originated from the work of the late Prof. P. J. Daniell in the early 1940's.

With reference to the third paragraph of Dr. Westcott's remarks, no simulator was involved, and the comparison between theoretical and experimental results taken from the actual non-linear system is given to emphasize the accuracy of the describing-function method used in the analysis.

The similarity of Mr. Grensted's equation to the one considered is only superficial; they represent systems having vastly different physical characteristics. From Section 7, our equation may be written in the form

$$\frac{d^2x}{dt^2} + \left(2c\frac{dx}{dt} + x\right)^n = 0$$

from which it is seen that both restoring force and damping force are raised to the nth power.

The method used in the paper by Drs. West and Douce is essentially a steady-state analysis, and additional information is required to assess the stability of the oscillation in a given case. This point has been covered in a further paper.*

We are very grateful to Prof. Tustin for his remarks concerning the describing-function technique, especially in contrast to the disparaging views of Dr. Westcott. For the purposes of mathematical analysis the use of the phase plane is limited to second-order systems, and further developments on these lines is unlikely. However, it is felt that it is an invaluable experimental technique giving insight into the physical behaviour and modes of operation. For the vast majority of systems, no matter the complexity, the step-function response is unique for a given amplitude of input and identical initial conditions, and this means that a complete phase-plane diagram can be constructed for systems of any order. At the University of Manchester, work has been proceeding fairly successfully with third- and fourth-order systems using the phase plane as the means of recording a large amount of data in a small space, and restricting the trajectories to those for step-function response with the initial values of the first and higher time derivatives zero. As for second-order systems, regions can be mapped out showing different operating modes.

The describing-function technique is by far the most promising and powerful analytical method yet developed. It is not limited to "small order" equations, but at its present stage it is limited to one non-linearity per loop and to frequency-independent non-linear characteristics. Neither of these two limitations are fundamental, and we are endeavouring to break them down; this is regarded as the most important part of our future programme.

We agree with Prof. Tustin that the task of evaluating the response to step-functions is more of academic interest than of engineering importance. However, its use in understanding the performance of a non-linear system and for assessing the system's adequacy for a given set of requirements is more nearly met by a series of transient responses as performance data than by a series of frequency-response curves. It is felt that the future trend will be towards analysing the response to random inputs of various nature.

The justification that Mr. Blake asks for—that the addition of a hard-spring characteristic is worth the extra complication—is given in the example described by Mr. Thompson. We are very sympathetic to Mr. Blake's views on "better response," and would add that high speeds and large changes of speed are bad also for "people carrying devices."

We must apologize to Mr. Thompson for upsetting his mechanical sensibilities. We have, however, sinned in good company.

* West, J. C., Douce, J. L., and Livesley, R. K.: "The Dual-Input Describing Function and its Use in the Analysis of Non-Linear Feedback Systems," *Proceedings I.E.E.*, Paper No. 1877 M, July, 1955 (103 B).

THE ADJUSTMENT OF CONTROL SYSTEMS FOR QUICK TRANSIENT RESPONS

By A. T. FULLER, M.A.

(*The paper was first received* 10*th July*, 1954, *and in revised form* 3*rd February*, 1955.)

SUMMARY

The time taken for correction of transient errors in linear feedback systems is examined theoretically. For any of a large class of such systems the quickest approach to complete correction is obtained when the system is adjusted to be quasi-critically damped. Furthermore, the time taken to achieve a large attenuation of an initial error is, in general, dependent on only one system parameter, termed the settling time-constant; and this is relatively easy to calculate when the system is quasi-critically damped.

These results, together with other known advantages of quasi-critical damping, suggest that the latter constitutes a useful general design criterion of stability.

LIST OF SYMBOLS

$A_1, A_2, \ldots A_k, A_l, \ldots A_n$ = Real parts of roots of characteristic equation.
$a_0, a_1, \ldots a_n$ = Coefficients in characteristic equation.
α = Constant defined in eqn. (1).
$B_1, B_2, \ldots B_k, B_l, \ldots B_n$ = Imaginary parts of roots of characteristic equation.
$C_0, C_1, \ldots C_m$ = Coefficients defined in eqn. (1).
D = Constant factor of C_m.
E = Error in controlled quantity.
E_T = Transient response of E.
E_1 = First undershoot in Fig. 2.
E_2 = First overshoot in Fig. 2.
$F_0, F_2, \ldots F_l$ = Coefficients defined in eqn. (17).
$f(p)$ = Left-hand side of characteristic equation.
$f'(p)$ = Derivative of $f(p)$ with respect to p.
g = Adjustable component in system.
M_1, M_2 = Determinants defined in eqn. (19).
μ = Loop gain.
μ_c = Value of loop gain which gives quasi-critical damping.
n = Order of characteristic equation.
ω = Angular frequency of harmonic factor of C_m.
p = Operator representing differentiation.
$Q_1, Q_2, \ldots Q_k, Q_l, \ldots Q_n$ = Roots of characteristic equation.
ρ = Loop integration rate.
ρ_c = Value of loop integration rate which gives quasi-critical damping.
s = Number of equal loop lags.
T = Loop lag in system of Fig. 1.
T_l = Loop delay.
$T_1, T_2, \ldots T_h$ = Loop lags.
t = Time.
t_s = Settling time.
τ_s = Settling time-constant.
τ_{sc} = Value of τ_s when system is quasi-critically damped.
θ = Constant part of phase of harmonic factor of C_m.
$u_0, u_1, \ldots u_n$ = Coefficients defined in eqn. (9).
$v_0, v_1, \ldots v_n$ = Coefficients defined in eqn. (9).

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Mr. Fuller is in the Department of Engineering, University of Cambridge, and was formerly with Waymouth Gauges and Instruments, Ltd.

(1) INTRODUCTION

One method of evaluating the performance of a control syste is through study of its response to a transient disturbanc Various theoretical criteria,[1] such as integral-squared-error, hav been proposed on this basis. The present paper describ another such criterion, the features of which are its mathematic simplicity and its practical appropriateness for certain types o control system.

If a control system, initially at rest, is subjected to a transien disturbance, the error in the controlled quantity may be large a first, but will eventually be reduced to comparatively sma proportions. The time taken to reach a given small proportio is a measure of the system performance. If the possibility o overshoot exists, a better measure is the time taken for the erro to reach a point after which it always remains within the give small proportion. This time may be termed[2] the *settling tim* of the process. The error will eventually approach some con stant value (which is often zero), and for this value the settlin time, being infinite, is no longer a useful measure of performance However, as will be shown, the ultimate approach is essentiall exponential and may thus be characterized by a time-constant which will be termed the *settling time-constant*. It follows that if a number of different systems are compared, that system which has the smallest settling time-constant will provide the quickes attenuation of transient error down to very small values. W shall accordingly adopt the settling time-constant as a criterio of performance, and investigate how it may be minimized b system adjustment.

The practical importance of the criterion may perhaps b judged by considering relevant examples as follows:

(*a*) A weapon-aiming servo mechanism which is suddenly switche from one target to another. A suitable criterion of performance i the time taken to become trained on the second target. Criteri involving weighted integrals of error are inappropriate, since al except very small errors are equally bad, i.e. "a miss is as good a a mile."

(*b*) An automatic pilot controlling an aircraft flying through bumpy weather. After a severe bump, passenger comfort may require that the aircraft should get back to steady flight on its original course as soon as possible.

(*c*) Processes in which fuel economy is important. It might be required that the system should attain its operating point as soon as possible after starting up.

We may anticipate results and add:

(*d*) Processes in which overshoot is inadmissible, e.g. bobbin-winding machines and temperature limiters in jet engines.

(*e*) Processes in which a conservative stability margin is required.

(*f*) Complex systems in the early theoretical stages of design when only the simplest criteria are mathematically manageable.

It should be noted, however, that the criterion may lead to an unduly conservative design in certain systems, such as those in which only a comparatively low attenuation of error is required, or those in which a high loop gain is more important than a fast response. Even so, the criterion may still be useful

for these systems in providing a simple indication of lower limits for design parameters.

(2) CALCULATION OF SETTLING TIME

It is well known[3,4] that the transient response of a stable linear system to an input may be expressed as the sum of a number of terms, each of which has a factor which decreases exponentially with time. Furthermore, it is easy to show that, as time progresses, the terms having the faster rates of decay become negligibly small compared with those having the least rate of decay.[2,4,5,6] Thus if a disturbance is applied to a linear control system, the transient response E_T, of the error E in the controlled quantity, for large values of time t, may be written in the following general form:

$$E_T = \varepsilon^{\alpha t}(C_0 + C_1 t + C_2 t^2 + \ldots + C_m t^m) \quad . \quad . \quad (1)$$

where α is a negative real constant, and $C_0, C_1, \ldots C_m$ either are real constants or vary sinusoidally with time. Consider first the case when the C's are constants. For sufficiently large values of t, i.e. for sufficiently large attenuation of error, eqn. (1) approaches

$$E_T = \varepsilon^{\alpha t} C_m t^m \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

From eqn. (2)

$$\log E_T = t\left(\alpha + \frac{\log C_m}{t} + \frac{m \log t}{t}\right) \quad . \quad . \quad . \quad (3)$$

When $t \to \infty$ the second and third terms in the bracket in eqn. (3) become negligibly small. Eqn. (3) then yields

$$t \to \alpha^{-1} \log E_T \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

From eqn. (4) the settling time t_s (as defined in Section 1), required for the error to reach a small value E_T, is given by

$$t_s \to \alpha^{-1} \log E_T \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Next consider the alternative case when $C_0, C_1, \ldots C_m$ vary sinusoidally with time. We may use the preceding argument, with C_m replaced by $D \sin(\omega t + \theta)$, where D, ω and θ are real constants. Thus the point when the system has settled to a given small error E_T is still given by eqn. (4), unless $\sin(\omega t + \theta)$ happens to be very small at this point, making $(\log C_m)/t$ no longer negligible in eqn. (3). But it may be shown as follows that the latter contingency does not arise. For sufficiently small deviations of time near $\sin(\omega t + \theta) = 0$, the graph of E_T against t may be approximated by a straight line crossing the t-axis. These values of E_T must therefore be exceeded in magnitude during subsequent values of t, i.e. the point at which $\sin(\omega t + \theta)$ equals zero must occur a finite time before the point at which the response settles to the given small error. Hence, at the latter point, $\sin(\omega t + \theta)$ cannot be small compared with $\varepsilon^{\alpha t}$ for sufficiently large values of t. Therefore eqn. (5) is a valid expression for settling time when $t \to \infty$.

Now α is a parameter either of the input disturbance or of the system. In the latter case it is the only system parameter which determines the long-term behaviour, according to eqn. (5), and so constitutes the criterion of performance. Minus the reciprocal of this system parameter will be called the settling time-constant, τ_s, of the system.

Thus $$\tau_s = -1/\alpha \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

If, on the other hand, α is a parameter of the input disturbance, eqn. (5) does not yield a criterion for the system, i.e. the disturbance is too slow fully to test the capabilities of the system. This case will not be considered further in the paper.

(3) MINIMIZATION OF SETTLING TIME

Let us vary some physical parameter of the system in an attempt to minimize the settling time-constant. Eqn. (6) shows that this implies minimization of α. Now α is the algebraically largest of the real parts of the roots of the characteristic equation of the system; this is clear from the methods[3,4] used in deriving eqn. (1). Therefore, as a first step, we calculate the rate of change of one of these roots with change of a general system parameter.

The characteristic equation of the system may be written $f(p) = 0$, in which

$$f(p) = a_0 + a_1 p + a_2 p^2 + \ldots + a_n p^n \quad . \quad . \quad (7)$$

where p is the operator representing differentiation and a_0, $a_1, \ldots a_n$ are constants readily calculable in terms of system parameters by means of conventional rules (e.g. Kirchhoff's laws in the case of electrical systems). If the roots of the characteristic equation are $Q_1, Q_2, \ldots, Q_n$,

$$f(p) = a_n(p - Q_1)(p - Q_2)\ldots(p - Q_n) \quad (8)$$

Suppose the physical component which is to be varied is g (this could represent an impedance, an inertia, a d.c. gain, or a rate of integration, etc.). It is well known in elementary network theory that coefficients such as $a_0, a_1, \ldots, a_n$ are linear functions of any single component, so that eqn. (7) may be written

$$f(p) = u_0 + u_1 p + \ldots + u_n p^n + g(v_0 + v_1 p + \ldots + v_n p^n) \quad (9)$$

where $u_0, u_1, \ldots, u_n$ and $v_0, v_1, \ldots, v_n$ are constants (some of which are usually zero).

Differentiating eqns. (8) and (9), which are identities in p, with respect to g we obtain

$$\begin{aligned} \frac{df}{dg} &= \frac{da_n}{dg}(p - Q_1)(p - Q_2)\ldots(p - Q_n) \\ &+ a_n\left(-\frac{dQ_1}{dg}\right)(p - Q_2)\ldots(p - Q_n) \\ &+ \ldots \\ &\vdots \\ &+ a_n(p - Q_1)(p - Q_2)\ldots\left(-\frac{dQ_n}{dg}\right) \quad . \quad . \quad (10) \end{aligned}$$

and $$\frac{df}{dg} = v_0 + v_1 p + \ldots + v_n p^n \quad . \quad . \quad . \quad . \quad (11)$$

Equating the right-hand sides of eqns. (10) and (11), and substituting $p = Q_1$, in order to eliminate

$$\frac{dQ_2}{dg}, \frac{dQ_3}{dg}, \text{etc.,}$$

yields

$$\frac{dQ_1}{dg} = -\frac{v_0 + v_1 Q_1 + v_2 Q_1^2 + \ldots + v_n Q_1^n}{a_n(Q_1 - Q_2)(Q_1 - Q_3)\ldots(Q_1 - Q_n)} \quad . \quad (12)$$

The interpretation of eqn. (12) is rendered less abstract if we first consider some special cases and then proceed to greater generality. The following notation will be used: the real and imaginary parts of $Q_1, Q_2, \ldots, Q_n$ will be $A_1, A_2, \ldots, A_n$ and $B_1, B_2, \ldots, B_n$ respectively. Further, it will be assumed that $A_1, A_2, \ldots, A_n$ are suffixed in order of magnitude,

i.e. $$0 > A_1 \geqslant A_2 \geqslant A_3 \ldots \geqslant A_n \quad . \quad . \quad . \quad . \quad (13)$$

Note that A_1 is thus minus the reciprocal of the settling time-constant,

i.e. $$A_1 = -1/\tau_s \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

Case (a).

$$v_1, v_2, \ldots, v_n \text{ are all zero}$$
$$Q_1, Q_2, \ldots, Q_n \text{ are all real}$$

This situation can arise, for example, when the open-loop transfer function has no zeros and only real, different poles, and when g represents loop gain.[6] Eqn. (12) becomes

$$\frac{dA_1}{dg} = -\frac{v_0}{a_n(A_1 - A_2)(A_1 - A_3)\ldots(A_1 - A_n)} \quad . \quad (15)$$

In this equation a_n is positive, since the system is being regarded as stable,[7] and the factors $(A_1 - A_2)$, etc., are also positive, from expressions (13). To simplify the argument we shall also consider v_0 to be positive, as is usually the case in practice, although it turns out that the sign of v_0 does not influence the final conclusions to be reached. Hence dA_1/dg is negative. Therefore the settling time-constant may be reduced, and the system performance improved, by increasing g.

This remains true as long as $Q_1, Q_2, \ldots, Q_n$ remain real. Actually, continued change of g causes members of a pair of adjacent roots to become first equal and then complex. In particular, considering Q_1 and Q_2, the sign of dA_2/dg may be shown to be positive by means of an expression analogous to eqn. (15). Hence increase of g causes A_1 and A_2 to approach one another, as well as improving performance.

Case (b).

The conditions are as in case (*a*), except that any roots, apart from Q_1, may be complex.

Suppose, for example, that Q_k is complex, Q_l is its conjugate, and the other Q's are real. Then the expression for dA_1/dg is the same as that given by eqn. (15) for case (*a*), except that the positive factors $(A_1 - A_k)$, $(A_1 - A_l)$, must be replaced by $(A_1 - A_k - jB_k)$, $(A_1 - A_l - jB_l)$. The product of the latter factors is

$$(A_1 - A_k)^2 + B_k^2$$

(putting $A_l = A_k$ and $B_l = -B_k$, since Q_k and Q_l are conjugate), which is necessarily positive. Hence the sign of dA_1/dg is the same as in case (*a*). The same argument may be applied to any further pairs of complex conjugate roots.

It follows that the results for case (*a*) are also valid for this more general case, so that the system may still be improved by increasing g, at least until A_1 becomes equal to A_2.

Case (c).

The conditions are as in case (*a*), except that Q_1 is complex, with Q_2 as its conjugate.

This situation arises when, in the system described under case (*a*), loop gain is increased beyond the point where A_1 becomes equal to A_2. A geometrical demonstration that Q_1 and Q_2 do, in fact, become complex beyond this point has been given in a previous paper.[6]

We have $$\frac{dA_1}{dg} = \text{Real part of } \frac{dQ_1}{dg} \quad . \quad . \quad . \quad . \quad (16)$$

Substituting in eqn. (16) an expression for dQ_1/dg derived from eqn. (12), noting that $A_2 = A_1$ and $B_2 = -B_1$ and rearranging, we find

$$\frac{dA_1}{dg} = \frac{v_0(F_0 - F_2B_1^2 + F_4B_1^4 - \ldots + F_iB_1^i)}{2a_n[(A_1 - A_3)^2 + B_1^2][(A_1 - A_4)^2 + B_1^2]\ldots[(A_1 - A_n)^2 + B_1^2]} \quad . \quad (17)$$

where $F_0, F_2, \ldots, F_i$ are constants given by

$F_0 =$ Sum of products of $(A_1 - A_3), (A_1 - A_4), \ldots, (A_1 - A_n)$ taken $(n - 3)$ at a time.
$F_2 =$ Sum of products taken $(n - 5)$ at a time.
$F_4 =$ Sum of products taken $(n - 7)$ at a time.
$= . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; .$
$F_i = \pm$ Sum of $(A_1 - A_3), (A_1 - A_4), \ldots, (A_1 - A_n)$ if n is even; or unity if n is odd.

In eqn. (17) the denominator is necessarily positive. Furthermore, since F_0 is also necessarily positive, the numerator is positive, provided that B_1 is small enough to render negligible the terms $-F_2B_1^2$, $-F_6B_1^6$, etc. Hence, with this proviso, dA_1/dg is positive, and the system may therefore be improved by decreasing g.

Case (d).

The conditions are as in case (*c*), except that any of the roots $Q_3, Q_4, \ldots, Q_n$ may be complex.

By considering one pair of complex conjugate roots at a time as was done in case (*b*), it may be shown that the result of case (*c*) is also valid for the present, more general, case. Hence, provided that B_1 is small, the system may be improved by decreasing g.

The implications of the results of cases (*a*) to (*d*) are as follows. If we commence with a system which is comparatively overdamped [case (*b*)], it may be improved by increasing the loop gain, μ, up to the point where the "dominant" pole, Q_1, becomes complex [case (*d*)]. At this point the rate of change of settling time-constant shows an abrupt change of sign, so that further increase of loop gain is detrimental. This effect is illustrated in Fig. 1, which is a plot of A_1 against μ for a single-loop control system containing three simple lags. Clearly the best performance occurs at the point when Q_1 becomes complex, although this point does not correspond to a "minimum" in the usual mathematical sense. This point is called a "sharp minimum" to distinguish it from the "smooth minimum" of calculus.

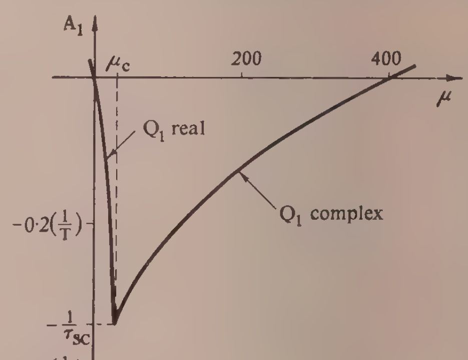


Fig. 1.—A_1 plotted against loop gain for a third-order system.

The system is a single loop comprised of three simple lags of values $200T$, T and T respectively; and a lagless amplifier of gain μ. The characteristic equation is thus

$$\mu + (1 + 200Tp)(1 + Tp)^2 = 0$$

A_1 is the largest of the real parts of the roots of this equation.

The conditions for the sharp minimum are also those for quasi-critical damping as may be seen from a definition of the latter:[6]

A system is quasi-critically damped when two or more of its poles have the same least value of decay rate, and one or more of these poles is real and negative.

Quasi-critical damping has been discussed in a previous paper,[6] from which it follows that a system in this condition has an interesting property. The response to a sudden change of input is such that "the output eventually reaches a regime in which it just fails to overshoot its final equilibrium value, this regime extending to infinite time." Thus systems which are just monotonic, or which are just free from overshoot, are often particular cases of quasi-critical damping.

Finally, if a completely general system is considered, with several non-zero v's, a study of eqn. (12) shows that usually quasi-critical damping still gives a sharp minimum. There are also a number of smooth maxima and minima, corresponding to zeros of the numerators in eqns. (12) and (17). For sufficiently complex systems, it is at least conceivable that the smooth minima could give better performance than the sharp minimum. However, for the simpler systems, quasi-critical damping represents an important minimum, and we confine attention to it in the paper.

The results of this Section confirm an assertion due to Thaler and Brown[8] to the effect that "critical damping" gives the quickest approach to "absolute correspondence."

(4) QUASI-CRITICAL DAMPING CALCULATIONS

It has been shown geometrically in the previous paper[6] that quasi-critical damping is a relatively easy condition to calculate, at least for a certain class of control systems. A somewhat more general approach will now be given, using methods from the theory of equations.

From the definition given in Section 3, it is clear that when $Q_1 = Q_2$ we have a particular type of quasi-critical damping. For the remainder of the paper, we shall confine attention to this type, which is relevant to many practical control systems.[6]

Thus quasi-critical damping is reached when two roots of $f(p)$ become equal, and the condition for such coincidence is known to be[9,10] as follows:

$$\begin{array}{l} \\ \\ \\ n-1 \text{ rows} \left\{ \begin{array}{l} \\ \\ \\ \\ \\ \\ \end{array} \right. \\ \\ n \text{ rows} \left\{ \begin{array}{l} \\ \\ \\ \\ \\ \\ \\ \end{array} \right. \end{array} \left| \begin{array}{cccccccccc} a_n & a_{n-1} & a_{n-2} & \cdots & a_1 & a_0 & 0 & 0 & \cdots & 0 \\ 0 & a_n & a_{n-1} & & a_2 & a_1 & a_0 & 0 & & 0 \\ 0 & 0 & a_n & & a_3 & a_2 & a_1 & a_0 & & 0 \\ \vdots & & & & & & & & & \vdots \\ 0 & & & & & & & & & a_0 \\ 0 & & & & & & & & & a_1 \\ \vdots & & & & & & & & & \vdots \\ 0 & 0 & na_n & & 3a_3 & 2a_2 & a_1 & 0 & & 0 \\ 0 & na_n & (n-1)a_{n-1} & & 2a_2 & a_1 & 0 & 0 & & 0 \\ na_n & (n-1)a_{n-1} & (n-2)a_{n-2} & \cdots & a_1 & 0 & 0 & 0 & \cdots & 0 \end{array} \right| = 0 \quad . \quad . \quad . \quad (18)$$

The condition for quasi-critical damping given by eqn. (18) is the same as that given in the previous paper,[6] except that, in the latter, the rows of the determinant were rearranged to bring out more clearly the analogy with one of the Routh–Hurwitz criteria.[7,11] (An interesting property of the above determinant is that its minors determine whether or not the system is "completely" aperiodic.[11])

A further result from the theory of equations is that, if an equation has two equal roots, their common value may be expressed by the ratio of two determinants involving the polynomial coefficients.[9,10] As applied to eqn. (7), this means that, if eqn. (18) is satisfied, the settling time-constant, termed τ_{sc} for this condition, is given by

$$\tau_{sc} = M_1/M_2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

where M_1 is the minor obtained from the determinant in eqn. (18) by deleting the first and last rows, and the first and last columns; and M_2 is the similar minor obtained by deleting the first and last rows, and the first and penultimate columns (an analogous expression exists[12] for the hunting frequency of a zero-damped system).

By way of illustration, for a third-order system, the condition for quasi-critical damping is, from eqn. (18),

$$\begin{vmatrix} a_3 & a_2 & a_1 & a_0 & \cdot \\ \cdot & a_3 & a_2 & a_1 & a_0 \\ \cdot & \cdot & 3a_3 & 2a_2 & a_1 \\ \cdot & 3a_3 & 2a_2 & a_1 & \cdot \\ 3a_3 & 2a_2 & a_1 & \cdot & \cdot \end{vmatrix} = 0 \quad . \quad . \quad . \quad . \quad . \quad (20)$$

If this is satisfied, the expression for the settling time-constant is

$$\tau_{sc} = \frac{\begin{vmatrix} a_3 & a_2 & a_1 \\ \cdot & 3a_3 & 2a_2 \\ 3a_3 & 2a_2 & a_1 \end{vmatrix}}{\begin{vmatrix} a_3 & a_2 & a_0 \\ \cdot & 3a_3 & a_1 \\ 3a_3 & 2a_2 & \cdot \end{vmatrix}} \quad . \quad . \quad . \quad . \quad . \quad (21)$$

$$= 2\frac{a_2^2 - 3a_1a_3}{a_1a_2 - 9a_0a_3} \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

When loop gain is the quantity which is to be varied to get quasi-critical damping, the process of solving eqns. (18) and (19) may often be replaced by the following quicker method.[6] Calculate the algebraically largest real root of $f'(p) = 0$, and substitute this value for p in $f(p) = 0$. The latter equation

will then yield an expression for μ_c, the loop gain which gives quasi-critical damping. Furthermore, the settling time-constant is minus the reciprocal of the above root, as is clear from the geometrical argument used in the previous paper.[6]

(5) EXAMPLES

The methods described in the last Section were used to derive the following results for some typical single-loop control systems. The values of μ_c have already been given in the previous paper,[6] but are included here for convenience:

System (a).

The dynamic loop elements are one large simple lag T_1, and one small simple lag T_2. Loop gain is given by

$$\mu_c = \tfrac{1}{4}[(T_1/T_2) - 2 + (T_2/T_1)] \quad . \quad . \quad . \quad (23)$$

The settling time-constant is given by

$$\tau_{sc} = 2[(1/T_2) + (1/T_1)]^{-1} \quad . \quad . \quad . \quad (24)$$

System (b).

One large simple lag T_1, and s small simple lags, each of value T_2:

$$\mu_c = \frac{T_1}{sT_2}\left[\left(1 - \frac{T_2}{T_1}\right)\Big/\left(1 + \frac{1}{s}\right)\right]^{s+1} \quad . \quad . \quad . \quad (25)$$

$$\tau_{sc} = \left(1 + \frac{1}{s}\right)\Big/\left(\frac{1}{sT_2} + \frac{1}{T_1}\right) \quad . \quad . \quad . \quad (26)$$

System (c).

One integrator, with rate of integration ρ, and one pure delay (or distance-velocity lag) of value T_l:

$$\rho_c = 1/(\varepsilon T_l) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

$$\tau_{cs} = T_l \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (28)$$

The advantages and limitations of quasi-critical damping are demonstrated for this system in Fig. 2, which shows the response of the error following a step-change of input (this may be readily plotted from data provided by Küpfmüller,[13] who gives an extensive treatment of the present system). Three different adjustments of the system are illustrated:

(i) $\rho = \tfrac{1}{2}\rho_c$, giving $\tau_s = 1\cdot 88 T_l$
(ii) $\rho = \rho_c$, giving $\tau_s = 1\cdot 00 T_l$ (i.e. quasi-critical damping)
(iii) $\rho = 2\rho_c$, giving $\tau_s = 4\cdot 31 T_l$

Examination of these graphs confirms that quasi-critical damping gives the quickest settling for small errors (i.e. less than $E_1 = 0\cdot 055$). On the other hand, the under-damped case (iii) gives the quickest settling for comparatively large errors (i.e. greater than $E_2 = 0\cdot 239$).

System (d).

One large simple lag T_1, and several small simple lags $T_2, T_3, \ldots, T_h$:

$$\mu_c \simeq (0\cdot 3)T_1/(T_2 + T_3 + \ldots + T_h) \quad . \quad . \quad (29)$$

to within about $\pm 25\%$.

$$\tau_{sc} \simeq 1\cdot 5(T_2 + T_3 + \ldots + T_h) \quad . \quad . \quad . \quad (30)$$

to within about $\pm 33\%$.

This last result is perhaps not obvious, but it may be derived by an argument similar to that used in the previous paper[6] for obtaining eqn. (29) above.

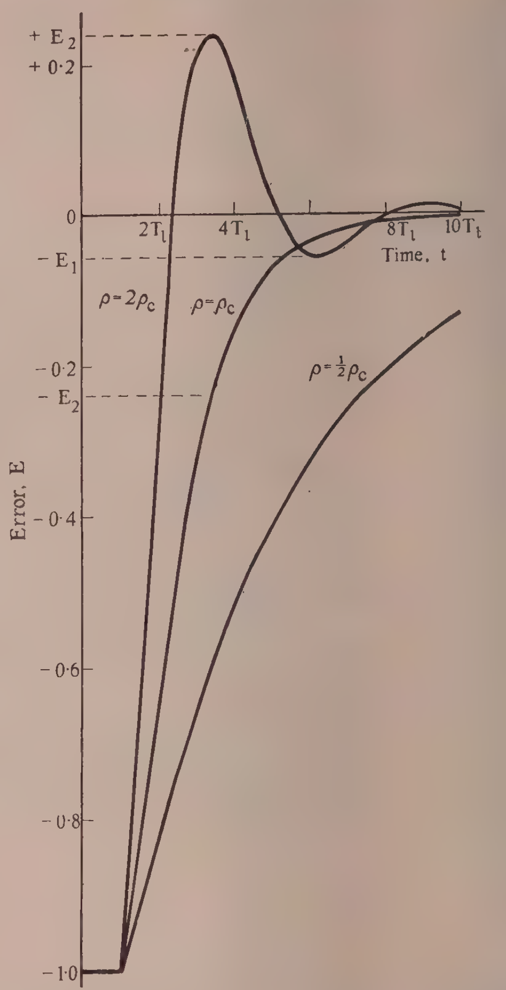


Fig. 2.—Responses of system (*c*), for three different adjustments of loop integration rate, to a unit-step input applied at $t = 0$.

The above results indicate that, for systems having only real lags distributed round the loop and having a reasonably large loop gain, the following conclusions may be drawn: the value of the minimum settling time-constant is of the order of the sum of the minor lags $T_2, T_3, \ldots, T_h$, i.e. of the effective "loop-delay" and is nearly independent of the major lag T_1, i.e. of the effective "loop-integration-rate."

(6) CONCLUSIONS

The time taken for any linear system to attenuate considerably the error due to any sudden transient disturbance is dependent on only one system parameter, namely the settling time-constant. This parameter is optimized, i.e. the quickest approach to complete correction is obtained, when the system is adjusted to be quasi-critically damped (at least for a large class of control systems). The adjustment for this condition is relatively simple to calculate, as is also the corresponding value of settling time-constant.

This favourable property of quasi-critical damping reinforces the argument for its use as a general design criterion of stability. It is especially appropriate for certain types of control system (as listed in Section 1).

(7) ACKNOWLEDGMENT

Thanks are due to the Directors of Waymouth Gauges and Instruments, Ltd., for permission to publish this paper.

(8) REFERENCES

(1) FICKEISEN, F. C., and STOUT, T. M.: "Analogue Methods for Optimum Servomechanism Design," *Applications and Industry* (American I.E.E.), 1952, No. 3, p. 244.

(2) CHU, Y.: "Synthesis of Feedback Control System by Phase-Angle Loci," *ibid.*, 1952, No. 3, p. 330.

(3) GARDNER, M. F., and BARNES, J. L.: "Transients in Linear Systems" (Wiley, New York, 1942), Volume I.

(4) CARSLAW, H. S., and JAEGER, J. C.: "Operational Methods in Applied Mathematics" (Oxford University Press, 1949), second edition, p. 280.

(5) MULLIGAN, J. H.: "The Effect of Pole and Zero Locations on the Transient Response of Linear Dynamic Systems," *Proceedings of the Institute of Radio Engineers*, 1949, **37**, p. 516.

(6) FULLER, A. T.: "Control Systems with Quasi-Critical Damping," *British Journal of Applied Physics*, 1954, **5**, p. 174.

(7) HURWITZ, A.: "Über die Bedingungen, unter welchen eine Gleichung nur Wurzeln mit negativen reellen Theilen besitzt," *Mathematische Annalen*, 1895, **46**, p. 273.

(8) THALER, G. J., and BROWN, R. G.: "Servomechanism Analysis" (McGraw-Hill, New York, 1953), p. 90.

(9) TRUDI, N.: "Teoria de' Determinanti" (Pellerano, Naples, 1862).

(10) MUIR, T.: "Theory of Determinants" (Macmillan, 1920), Volume III, p. 347.

(11) FULLER, A. T.: "Conditions for Aperiodicity in Linear Systems," *British Journal of Applied Physics*, 1955, **6**, p. 195.

(12) FRAZER, R. A., and DUNCAN, W. J.: "On the Criteria for the Stability of Small Motions," *Proceedings of the Royal Society*, 1929, **124A**, p. 642.

(13) KÜPFMÜLLER, K.: "Die Vorgänge in Regelsystemen mit Laufzeit," *Archiv der Elektrischen Ubertragung*, 1953, **7**, p. 71.

621–526

Paper No. 186
Sept. 1

THE INTEGRAL-OF-ERROR-SQUARED CRITERION FOR SERVO MECHANISMS

By H. H. ROSENBROCK, B.Sc.(Eng.), Ph.D., Associate Member.

(*The paper was first received* 15*th January, and in revised form* 18*th March*, 1955.)

SUMMARY

A graphical method is given for obtaining the integral of error-squared for a servo mechanism supplied with a given input signal, and the usefulness of this criterion is discussed. The servo mechanism is characterized by its open-loop response locus, and the input signal by its power spectrum. The method is extended to allow for noise at the input.

(1) INTRODUCTION

It has been suggested[1] that the integral of error-squared,

$$S = \int_0^\infty \epsilon^2(t)dt \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

would form a simple and convenient criterion of performance for a servo mechanism. Under certain conditions which are usually fulfilled S can be obtained from information about the error as a function of frequency; in fact,[2]

$$S = \frac{1}{\pi}\int_0^\infty |\epsilon(\omega)|^2 d\omega \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Moreover, for a linear system having an open-loop transfer function $G(p)$ we have

$$|\epsilon(\omega)|^2 = \frac{|\theta_i(\omega)|^2}{|1 + G(\omega)|^2} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where $|\theta_i(\omega)|^2$ is proportional to the quadratic spectrum of the input. Thus the input signal may be characterized by its frequency spectrum alone, and in simple cases S can be evaluated analytically from eqns. (2) and (3).

In those cases where S has been calculated, a step-function has often been used for the input θ_i. It is then found that the system which gives the least value of S has too little damping to be satisfactory in practice. Westcott,[3] following Hall,[1] has suggested that this is because the criterion penalizes a system for failing to respond immediately after a step change of input, which the system is physically incapable of doing. This is equivalent to saying that a servo mechanism cannot respond to arbitrarily high frequencies, and should not be penalized for failing to do so. In place of S, Westcott suggests as a criterion W, where

$$W = \int_0^\infty t\epsilon^2(t)dt \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

A graphical method will now be given by which S may be found for a given system supplied with a given input signal. The system may be characterized by its open-loop harmonic-response locus, and the input by its power spectrum. Some examples will be worked out, and then the argument against S as a criterion of performance will be re-examined.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Dr. Rosenbrock is with Costain–John Brown Ltd.

(2) GRAPHICAL METHOD

From eqns. (2) and (3) we have

$$S = \frac{1}{\pi}\int_0^\infty \frac{|\theta_i(\omega)|^2}{|1 + G(\omega)|^2} d\omega \quad . \quad . \quad . \quad .$$

If now we plot

$$y = |1 + G(\omega)|^{-2} \quad . \quad . \quad . \quad . \quad .$$

against

$$x = \int_{\omega_0}^\omega |\theta_i(\omega)|^2 d\omega \quad . \quad . \quad . \quad . \quad .$$

we have

$$\int_{\omega=0}^\infty y dx = \int_0^\infty y \frac{dx}{d\omega} d\omega$$

$$= \int_0^\infty \frac{|\theta_i(\omega)|^2}{|1 + G(\omega)|^2} d\omega$$

$$= \pi S \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad .$$

When the signal θ_i has finite energy, ω_0 may be zero, and it convenient to use the dimensionless quantity P instead of where

$$P = \frac{\int_0^\infty \frac{|\theta_i(\omega)|^2}{|1 + G(\omega)|^2} d\omega}{\int_0^\infty |\theta_i(\omega)|^2 d\omega} \quad . \quad . \quad . \quad . \quad .$$

In Figs. 3 and 4, P is given by the ratio of the area under t curve $y = |1 + G|^{-2}$ to that under the line $y = 1$. For su input signals as an infinitely-long step the integral (7) diverges ω_0 tends to zero, and the base-line of x is infinitely long, as Figs. 5 and 6.

The scale of x depends only on the power spectrum of t input and not on the properties of the servo mechanism. T scale of y is invariable. Hence charts can easily be prepar either for a given input characterized by a measured spectrur or for inputs considered typical under given conditions.

The value of $|1 + G|$ is readily available from the open-loc harmonic-response locus of the system, and is merely the distan from the locus to the point $(-1, 0)$. The frequency response the system need not be given analytically, but may be measure The effect upon P of a change in the open-loop gain is particular easy to find, since it is only necessary to choose a new point t be $(-1, 0)$, change the scale of the response locus appropriatel and plot the new values of $|1 + G|$ against the same scale of .

Besides giving the value of P (or S), the graphical methc indicates the frequencies at which the greatest contributions t P (or S) arise. It therefore shows the frequencies at which a improvement of the system will be most profitable.

(3) EXAMPLES

The servo mechanism which will be considered is that treated Westcott[3] and shown in Fig. 1. It consists of a Ward Leonard

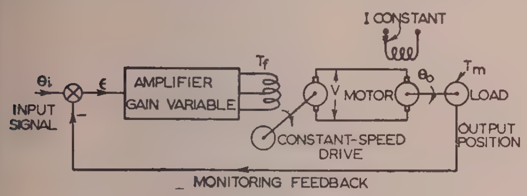


Fig. 1.—Servo system used as an example.

driving an inertia load, the generator having a split field fed ferentially by a valve amplifier of adjustable gain. We have

$$G(\omega) = \frac{K_v}{-j\omega^3 T_m T_f - \omega^2(T_m + T_f) + j\omega} \quad . \quad . \quad (10)$$

here $K_v = K_1K_2$ = Velocity-error constant.
K_1 = Amplifier output voltage per unit error angle, volts/rad.
K_2 = Velocity-constant of the motor, rad/sec/volt.
T_m = Mechanical time-constant of the motor and its load, sec.
T_f = Time-constant of the generator field, sec.

e shall put $T_m = 5$ sec and $T_f = 1$ sec, and shall treat K_v the design parameter. Westcott has shown analytically that r a step-function input S is least for $K_v = 0\cdot33$, while W is ast for $K_v = 0\cdot21$. This last setting gives a transient response hich would be accepted as satisfactory on the basis of practical perience.

The first example uses an input having a power spectrum oportional to $\varepsilon^{-k\omega}$. The form of the spectrum is shown in rve (i) of Fig. 2, with four scales (a), (b), (c) and (d), corre-

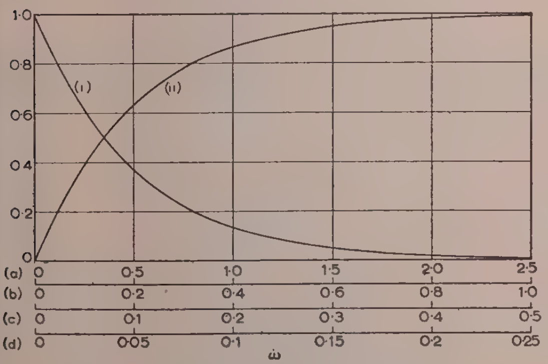


Fig. 2.—Power spectrum of the input signal.

Curve (i) shows the form of the assumed power spectrum, which is proportional $\varepsilon^{-k\omega}$. The four scales (a), (b), (c) and (d) are drawn respectively for $k=2$, 5, 10 d 20. Curve (ii) gives kx where $x = \int_0^{\omega} \varepsilon^{-k\omega} d\omega$. The resonant frequency of the vo system is near $\omega = 0\cdot25$.

onding respectively to $k = 2$, 5, 10 and 20. Increase of k om k_1 to k_2 is equivalent in its effect on P to reducing all the me-constants of the servo mechanism in the ratio k_1/k_2. urve (ii) of Fig. 2 shows kx, where $x = \int_0^{\omega} \varepsilon^{-k\omega} d\omega$.

Fig. 3 shows the error chart for an input having its power spec-um proportional to $\varepsilon^{-2\omega}$ and curve (a) of Fig. 8 shows the

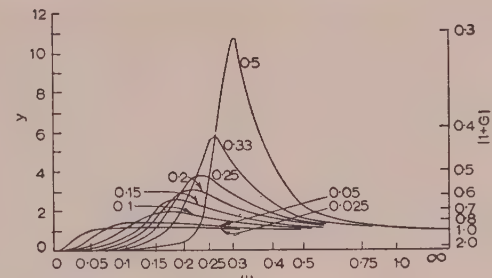


Fig. 3.—Error chart for the system of Fig. 1 when the input signal has a power spectrum proportional to $\varepsilon^{-2\omega}$.

variation of P with K_v. The criterion here calls for very high damping. Fig. 3 represents a case which is likely to occur only if the system is required to give a smoothed representation of the input.[4] In such a case very high damping would, in fact, be used. Although the least value of P is nearly unity, Fig. 3 shows that the low-frequency components of the input are accurately followed.

Curve (b) of Fig. 8 is drawn for the input represented by scale (b) of Fig. 2. To the accuracy of drawing and measurement used, the minimum value of P has the same value as in the last example and occurs at the same value of K_v. The only advantage derived from the increased relative value of the servo-mechanism resonant frequency seems to be a reduction in the penalty for using too large a value of K_v.

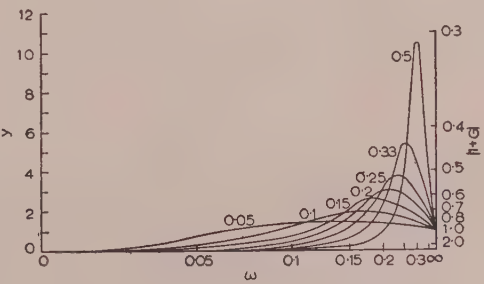


Fig. 4.—Error chart for the system of Fig. 1 when the input signal has a power spectrum proportional to $\varepsilon^{-10\omega}$.

Fig. 4 and curve (c) of Fig. 8 show a more practical case. The servo-mechanism resonance is seen from scale (c) of Fig. 2 to lie above the frequencies carrying the greater part of the input energy. The criterion now calls for a value of K_v of about $0\cdot4$, but shows that little penalty is incurred by using $K_v = 0\cdot21$ in order to give a better margin of stability. The least value of P is about $0\cdot64$, which shows that the relative increase of the servo-mechanism resonant frequency has produced a great improvement in performance. The important parameter is no longer K_v, as it was in the two previous examples, but it is now the speed of response.

Curve (d) of Fig. 8 shows conditions when the servo resonance is at a relatively high frequency. The criterion now calls for very light damping, but it will be shown later that the presence of noise in the input signal modifies this conclusion.

Fig. 5 is drawn for a step-function input, and curve (e) of Fig. 8 shows the corresponding values of S. These values are calculated, and are shown to an arbitrary scale since the signal energy is now infinite. Fig. 6 and curve (f) of Fig. 8 correspond to an input proportional to $1 + 10\varepsilon^{-t}$. The least value of S occurs for $K_v \simeq 0\cdot05$.

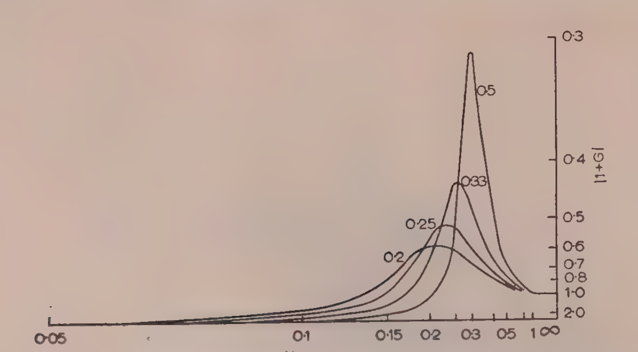


Fig. 5.—Error chart for the system of Fig. 1 when the input signal is a step-function.

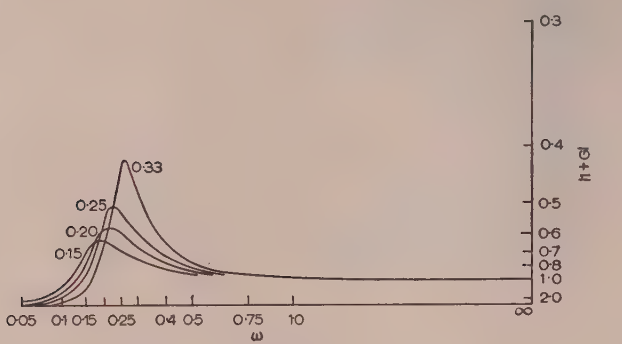


Fig. 6.—Error chart for the system of Fig. 1 when the input signal is proportional to $1 + 10\varepsilon^{-t}$.

(4) EFFECT OF NOISE

By an extension of the graphical method given in Section 2 we may take account of noise. If the signal θ_i is applied to the input together with noise θ_{ni}, the output consists of a signal θ_0 together with noise θ_{n0}. The difference between the output and θ_i is

$$\delta = \theta_0 + \theta_{n0} - \theta_i = -\epsilon + \theta_{n0} \quad . \quad . \quad . \quad (11)$$

where ϵ as before means $\theta_i - \theta_0$. We therefore take as our criterion

$$T = \int_0^\infty \delta^2(t)dt$$

$$= \int_0^\infty [\epsilon(t) - \theta_{n0}(t)]^2 dt$$

$$= \int_0^\infty \epsilon^2(t)dt + \int_0^\infty \theta_{n0}^2(t)dt - 2\int_0^\infty \epsilon(t)\theta_{n0}(t)dt \quad . \quad (12)$$

We now suppose that θ_i and θ_{ni} are representative samples of a statistically uniform signal and its accompanying noise. If then the noise is uncorrelated with the signal, and if the samples are sufficiently long, it is physically evident that the last integral in eqn. (12) will be negligible in comparison with the first two. This can be proved from the mathematical definition of correlation,[2] although the proof is not simple.

Thus
$$T = S + \int_0^\infty \theta_{n0}^2(t)dt$$

$$= S + \frac{1}{\pi}\int_0^\infty |\theta_{n0}(\omega)|^2 d\omega$$

$$= S + \frac{1}{\pi}\int_0^\infty \left|\frac{\theta_{ni}(\omega)G(\omega)}{1 + G(\omega)}\right|^2 d\omega \quad . \quad . \quad . \quad (1$$

and we find the value of the last integral by plotting

$$y = \left|\frac{G(\omega)}{1 + G(\omega)}\right|^2 \quad . \quad . \quad . \quad . \quad . \quad (1$$

against
$$x = \int_0^\omega |\theta_{ni}(\omega)|^2 d\omega \quad . \quad . \quad . \quad . \quad . \quad (1$$

As before it is convenient to divide throughout by $\frac{1}{\pi}\int_0^\infty |\theta_i(\omega)|^2 d$ thus obtaining

$$Q = P + \frac{\int_0^\infty \left|\frac{\theta_{ni}(\omega)G(\omega)}{1 + G(\omega)}\right|^2 d\omega}{\int_0^{\omega_1} |\theta_{ni}(\omega)|^2 d\omega} \cdot \frac{\int_0^{\omega_1} |\theta_{ni}(\omega)|^2 d\omega}{\int_0^\infty |\theta_i(\omega)|^2 d\omega} \quad . \quad (1$$

$$= P + N, \text{ say} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1$$

The reason for writing N in the form given in eqn. (16) is that is then a product of two dimensionless ratios, one of areas an one of energies, and no difficulty arises over scale factors. An convenient value can be chosen for ω_1, e.g. that value whic makes

$$\int_0^{\omega_1} |\theta_{ni}(\omega)|^2 d\omega = \int_0^\infty |\theta_i(\omega)|^2 d\omega \quad . \quad . \quad . \quad (1$$

Fig. 7 shows noise charts plotted from eqns. (14) and (15) fo the system of Fig. 1 when white noise is applied to the inpu The variation of N with K_v is shown by curve (g) of Fig. 8, th

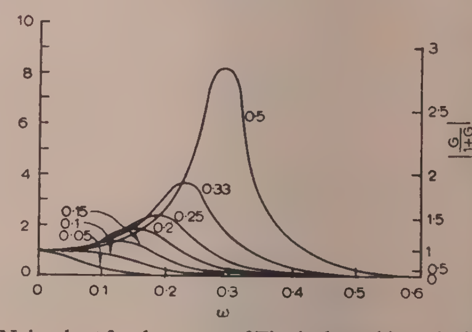


Fig. 7.—Noise chart for the system of Fig. 1 when white noise is applie to the input.

scale of which is such that the noise energy in the band $\omega =$ to $\omega = 1$ is equal to the energy in the signals for which P shown in the same Figure. Curves (c_1) and (c_2) show respectivel the effect of adding 10% and 20% of N to curve (c), whic relates to Fig. 4. Thus curve (c_2), for example, shows the valu of Q for the system of Fig. 1 when the spectrum of the signal as shown by scale (c) of Fig. 2 and white noise is present with th

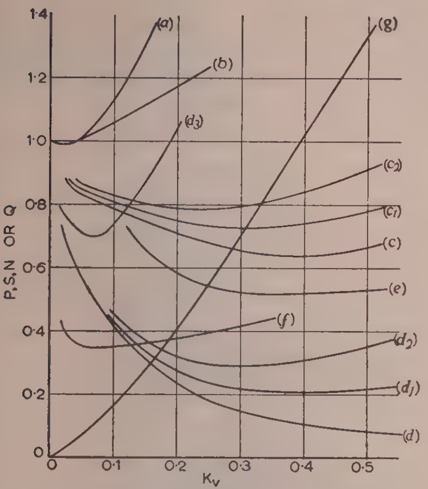


Fig. 8.—Variation of P, S, N or Q with K_v.

The curves all relate to the system of Fig. 1 and show the following:

(a) P for signal with power spectrum proportional to $\varepsilon^{-2\omega}$.
(b) P for signal with power spectrum proportional to $\varepsilon^{-5\omega}$.
(c) P for signal with power spectrum proportional to $\varepsilon^{-10\omega}$.
(d) P for signal with power spectrum proportional to $\varepsilon^{-20\omega}$.
(e) S for step-function input (scale arbitrary).
(f) S for input proportional to $1 + 10\varepsilon^{-t}$ (scale arbitrary).
(g) N for white noise at input having energy in the band $\omega = 0$ to $\omega = 1$ equal to that of the input signals for (a), (b), (c) or (d).

(c_1) Q for (c) with 10% of (g).
(c_2) Q for (c) with 20% of (g).
(d_1) Q for (d) with 10% of (g).
(d_2) Q for (d) with 20% of (g).
(d_3) Q for (d) with 200% of (g).

gnal and has the same energy in the band $\omega = 0$ to $\omega = 5$ as e total energy of the signal. Curves (d_1), (d_2) and (d_3) of g. 8 show respectively the effects of adding 10, 20 and 200% N to curve (d). Curves (c_1), (c_2), (d_1), (d_2) and (d_3) illustrate e fact that the presence of noise requires greater damping.

(5) DISCUSSION OF THE EXAMPLES

From Figs. 3 and 4 it is evident that the important factor in eciding the value of K_v for minimum P is the amount of energy the signal at frequencies near the servo resonance. The iterion will require high damping if the servo resonance is ithin a frequency band where the input has appreciable energy. the servo resonance is not within such a band, the criterion lows lower damping than experience shows to be desirable. he last result is not surprising, since by using P we ignore the fects of noise, and since the practical setting of a servo system ust allow for possible changes in the parameters, which with ry small damping might lead to instability.

The transition from conditions which demand high damping those which allow light damping is more sudden than might expected, and it is accompanied by a sharp decrease in the ast value of P. This result may be a consequence of the parular choice of input signal spectrum.

Returning to the criticisms which have been made of the integral error-squared as a performance criterion, we see from Fig. 5 at, with a step-function input, the criterion allows light damping cause there is relatively little energy in the signal at the servo sonant frequency. The high-frequency region where G is gligible contributes to S a quantity which is nearly independent K_v, and has little effect in locating the minimum of S. This is early illustrated also in Fig. 6, where the system is penalized ore severely for its failure to respond to a sudden change of put, and yet the criterion requires greater damping owing to e greater energy in the signal near the servo resonant frequency.

It is shown in Appendix 10 that the use of W is equivalent to retaining S as the criterion but altering the input signal. The modified input has, in fact, a smaller proportion of the highest frequencies, but this is of secondary importance. A more important effect of the modification is to introduce into the input a component at each of the (complex) natural frequencies of the system. It is for this reason that the use of W leads to more highly damped systems than are obtained with S.

There seems to be a confusion of purpose in applying S as a criterion with a step-function input. We ought probably to distinguish two sets of conditions which a servo mechanism must satisfy:

(a) The system must be stable, and it must have a sufficient margin of stability to cover variations of the parameters which may occur when it is in service. The practical criteria of phase margin, maximum modulus, step-function response, etc., all relate to the margin of stability.* They can all be applied regardless of the signals which the servo mechanism will in fact be called on to follow.

(b) The output must fit as closely as possible, according to some accepted performance criterion, to the input. The integral of error-squared is one of many possible measures of this closeness of fit. All such criteria will, in general, give results which depend on the character of the input signal and of any noise which may be present. All will be overridden if they suggest less damping than the stability criteria require.

It may well be that, in many practical cases, the least value of Q occurs for rather less damping than is required by the stability criteria, but Q is little increased by the necessary increase of damping. This would explain why the stability criteria form a useful guide to the best practical setting. Thus it may be only in occasional cases that we need to use Q to fix the damping (Fig. 8). A performance criterion may, on the other hand, determine parameters about which the stability criteria give no information. It may also be useful in deciding whether a proposed alteration of the system is worth making, or in drawing up a specification of performance to be fulfilled by the servo mechanism when supplied with a given input.

When it is used as a specification of performance, the restriction on P can be supplemented by limits on the contribution to P which may arise in given frequency bands. Testing an existing servo mechanism to see whether it conforms to such a specification requires only the measurement of the amplitude of error for harmonic inputs at appropriate frequencies. No measurements of phase are required. The use of Q for purposes of specification requires a little more care since, for a given input signal and given noise, there is an absolute minimum below which Q cannot be reduced.[2]

(6) APPLICATION TO PROCESS CONTROL

The typical problem of a single-loop process-control system is shown in Fig. 9. Here A, B and C represent the fixed transfer

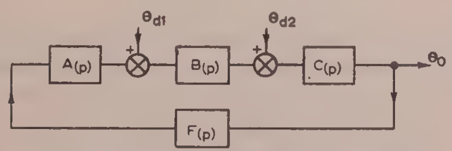


Fig. 9.—Process-control loop with disturbances.
ABCF = G

functions of parts of the process, F is the transfer function of the controller, θ_{d1} and θ_{d2} are disturbances, and θ_i, the set value, remains constant and is therefore made equal to zero. There may be more than two disturbances such as θ_{d1} and θ_{d2}. The object is to reduce the output variations θ_0, and if these may be measured adequately by their mean-square value we may apply the graphical method already given.

* The accepted numerical values used with these criteria may, of course, have some regard to performance under typical conditions, as well as to stability.

We first transfer all the disturbances to the output, obtaining a disturbance θ_d which is equivalent to θ_{d1}, θ_{d2}, etc., and can be calculated from them. It is likely, however, that θ_d will be obtained by measurement, since it is equal to θ_0 when the feedback loop is broken. Thus the quadratic spectrum of θ_d can be obtained by measuring that of θ_0 when $F = 0$.

When the loop is closed we have

$$\int_0^\infty \theta_0^2(t)dt = \frac{1}{\pi}\int_0^\infty |\theta_0(\omega)|^2 d\omega \quad . \quad . \quad . \quad . \quad (19)$$

$$= \frac{1}{\pi}\int_0^\infty \frac{|\theta_d(\omega)|^2}{|1 + G(\omega)|^2} d\omega \quad . \quad . \quad . \quad (20)$$

We therefore take as the criterion of performance

$$D = \frac{\displaystyle\int_0^\infty \frac{|\theta_d(\omega)|^2}{|1 + G(\omega)|^2} d\omega}{\displaystyle\int_0^\infty |\theta_d(\omega)|^2 d\omega} \quad . \quad . \quad . \quad . \quad (21)$$

which is similar in form to P, and may be found in the same way by plotting

$$y = |1 + G(\omega)|^{-2} \quad . \quad . \quad . \quad . \quad . \quad (22)$$

against

$$x = \int_0^\omega |\theta_d(\omega)|^2 d\omega \quad . \quad . \quad . \quad . \quad . \quad (23)$$

We then have to minimize D by adjusting the parameters of F, confirming afterwards that the setting so obtained gives an adequate margin of stability or, if not, making the necessary changes. It will be noticed that this analysis corresponds to that given in Section 2 for a servo mechanism in the absence of noise.

(7) CONCLUSIONS

The graphical method which has been given allows the integral of error-squared to be obtained without too much labour even for moderately complicated systems. It is suggested that when this integral is used as a performance criterion it should be evaluated for the type of input to which the system must respond (including noise, if this is appreciable) and that separate consideration should be given to the margin of stability. Under these conditions the evidence[2] seems to indicate that the criterion will give results in accordance with practical judgment.

(8) ACKNOWLEDGMENTS

The author wishes to express his thanks to Dr. J. H. Westcott for the stimulating discussion which preceded the writing of this paper, and to the Directors of Costain–John Brown Ltd. for permission to publish the paper.

(9) REFERENCES

(1) HALL, A. C.: "The Analysis and Synthesis of Linear Servomechanisms" (Technology Press, Massachusetts Institute of Technology, 1943).

(2) JAMES, H. M., NICHOLLS, N. B., and PHILLIPS, R. S.: "Theory of Servomechanisms," Radiation Laboratory Series No. 25 (McGraw-Hill, 1947), Chapters 6 and 7.

(3) WESTCOTT, J. H.: "The Minimum-Moment-of-Error-Squared Criterion: A New Performance Criterion for Servo Mechanisms," *Proceedings I.E.E.*, Paper No. 1644 M, March, 1954 (**101**, Part II, p. 471).

(4) BELL, J., AND LANGHAM, E. M.: "Mean-Wind-Veloc Instruments for use on Naval Craft," *ibid.*, Pa No. 1103 M, January, 1951 (**98**, Part II, p. 448).

(5) COPSON, E. T.: "Theory of Functions of a Complex Variabl (Oxford University Press, 1935).

(10) APPENDIX

We define

$$W_n[g(p)] = \int_0^\infty t^n \epsilon_g^2(t)dt \quad . \quad . \quad . \quad . \quad (2$$

where $\epsilon_g(t)$ is the error for an input $g(t)$. Then we have[3]

$$W_n[g(p)] = (-)^n \lim_{\sigma \to 0}\left[\frac{\partial^n}{\partial\sigma^n}\int_0^\infty \epsilon_g^2(t)\epsilon^{-\sigma t}dt\right] \quad . \quad . \quad . \quad (2$$

$$= (-)^n \lim_{\sigma \to 0}\left[\frac{\partial^n}{\partial\sigma^n}\frac{1}{2\pi j}\int_{-j\infty}^{j\infty} \epsilon_g(p)\epsilon_g(\sigma - p)dp\right] . \quad (2$$

provided that the poles of $\epsilon_g(p)$ are all in the left-hand half-pla Now σ is initially real, but we may regard the integral in eqn. (2 as defining a function $F(\sigma)$ of the complex variable σ. Then f σ within a small region about the origin we can show[5] that $F($ is an analytic function with derivatives of all orders which m be found by differentiating under the sign of integration. follows that

$$W_n[g(p)] = (-)^n \lim_{\sigma \to 0}\left[\frac{1}{2\pi j}\int_{-j\infty}^{j\infty} \epsilon_g(p)\epsilon_g^{(n)}(\sigma - p)dp\right] \quad . \quad . \quad (2$$

$$= (-)^n \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} \epsilon_g(p)\epsilon_g^{(n)}(-p)dp \quad . \quad . \quad . \quad . \quad . \quad . \quad (2$$

$$= (-)^n \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} \epsilon_g(-p)\epsilon_g^{(n)}(p)dp \quad . \quad . \quad . \quad . \quad . \quad . \quad (2$$

$$= \frac{1}{2}(-)^n \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} [\epsilon_g(p)\epsilon_g^{(n)}(-p) + \epsilon_g(-p)\epsilon_g^{(n)}(p)]dp \quad . \quad (3$$

$$= \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} \epsilon_g(p)\epsilon_g(-p)H(p)dp, \text{ say} \quad . \quad . \quad . \quad . \quad . \quad (3$$

where

$$H(p) = (-)^n \frac{1}{2}\left[\frac{\epsilon_g^{(n)}(p)}{\epsilon_g(p)} + \frac{\epsilon_g^{(n)}(-p)}{\epsilon_g(-p)}\right] \quad . \quad . \quad . \quad (3$$

But $H(p) = H(-p)$, and we may therefore write

$$H(p) = h(p)h(-p) \quad . \quad . \quad . \quad . \quad . \quad . \quad (3$$

where all the poles of $h(p)$ are in the left-hand half-plane, provid that $\epsilon_g(p)$ has no purely imaginary zeros. In general $h(p)h(-$ has poles at the poles and zeros of $\epsilon_g(p)\epsilon_g(-p)$.

Then

$$W_n[g(p)] = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} \epsilon_g(p)h(p)\epsilon_g(-p)h(-p)dp \quad . \quad (3$$

$$= W_0[g(p)h(p)] \quad . \quad . \quad . \quad . \quad . \quad . \quad (3$$

The function $h(p)$ will not generally be unique, as there will l some choice in the location of its zeros.

It is evident from eqns. (32) and (33) that $|g(\omega)h(\omega)|$ will l less than $|g(\omega)|$ at very high frequencies, but that it may l considerably greater near the resonant frequency of the syster

ne function $g(p)h(p)$ has a pole at each complex natural freuency of the system, and if there is a lightly-damped mode the ne-function corresponding to $g(p)h(p)$ will be oscillatory with e corresponding frequency.

For example, if

$$g(p) = \frac{1}{p} \text{ and } \epsilon_g(p) = \frac{p}{p^2 + \omega p + \omega^2} \quad . \quad . \quad (36)$$

e have

$$-\frac{1}{2}\left[\frac{\epsilon_g'(p)}{\epsilon_g(p)} + \frac{\epsilon_g'(-p)}{\epsilon_g(-p)}\right]$$

$$= \frac{\omega(\omega + p)(\omega - p)}{(p^2 + \omega p + \omega^2)(p^2 - \omega p + \omega^2)} \quad . \quad . \quad (37)$$

d we may take

$$h(p) = \frac{(\omega + p)\sqrt{\omega}}{p^2 + \omega p + \omega^2} \quad . \quad . \quad . \quad . \quad (38)$$

or

$$h(p) = \frac{(\omega - p)\sqrt{\omega}}{p^2 + \omega p + \omega^2} \quad . \quad . \quad . \quad . \quad (39)$$

The time-function corresponding to $g(p)h(p)$ is then

$$\frac{1}{\sqrt{\omega}}\left[1 - \varepsilon^{-\omega t/2}\left(\cos\frac{\sqrt{(3)}\omega t}{2} - \frac{1}{\sqrt{3}}\sin\frac{\sqrt{(3)}\omega t}{2}\right)\right] . \quad (40)$$

or

$$\frac{1}{\sqrt{\omega}}\left[1 - \varepsilon^{-\omega t/2}\left(\cos\frac{\sqrt{(3)}\omega t}{2} + \sqrt{3}\sin\frac{\sqrt{(3)}\omega t}{2}\right)\right] . \quad (41)$$

Thus for the system of this example the value of $W \equiv W_1$ for a unit step-function input is the same as $S \equiv W_0$ when the input is given by eqn. (40) or (41). By using the standard forms given in Reference 3 it is easy to verify this result, and the common value of W and S is found to be $1/2\omega^2$.

DISCUSSION ON "THE MANCHESTER–KIRK O'SHOTTS TELEVISION RADIO-RELAY SYSTEM"*

WESTERN CENTRE, AT CARDIFF, 1ST NOVEMBER, 1954

SOUTHERN CENTRE, AT PORTSMOUTH, 9TH MARCH, 1955

Mr. J. S. Whyte (*at Cardiff*): On the T.D.2 relay system perating in the United States the required amplification at per-frequency is provided by triodes operating between itable resonant cavities; the output amplifier has a gain of 18 dB hen the output power is 0·5 watt. This arrangement has the lvantage that an h.t. supply of only a few hundred volts is quired compared with the 3 kV stabilized supply demanded the travelling-wave-valve amplifier and some 60 watts of power quired by its focusing coil. What are the authors' views on e relative merits of the two methods? Is it not a fact that e inherent broad-band characteristic of the travelling-wave nplifier, so often quoted as a great advantage, is reduced to figure not greatly exceeding that of the T.D.2 amplifiers, by e input and output matching difficulties? Would the authors ve used the travelling-wave-valve amplifier if suitable triodes d been available in this country when the system was designed?

Is there a routine for checking the automatic-starting arrangeents for the Diesel generators, and if so, how frequently do ey fail to start without necessarily interrupting traffic?

Mr. G. D. Curtis (*at Cardiff*): It would appear that the use supervisory equipment for the location of faults in unattended ephone radio-relay stations emanates from the desire to reduce e maintenance manpower required at such stations.

I assume that selected maintenance men in the area are called t in the event of a fault on such a relay station. In view of e complexity and additional fault liability incurred by the use supervisory fault detection, has there in practice been any ving in the cost of running such stations, or does the interest the capital cost of the supervisory equipment plus its deprecian total more than the calculated cost of manning such stations ring working hours?

The authors have stated that the group-delay/frequency aracteristics of the radio system vary with frequency from proximately 50 to 90 millimicrosec. I assume that this delay ows up as a phase displacement varying with frequency.

Would the authors give a comparison between the group-delay aracteristics of a radio system and those which one would expect en using coaxial cable?

Mr. W. P. Warren (*at Cardiff*): Could the authors give some indication of the electrical loading of the repeater stations used on the Manchester–Kirk o'Shotts radio-relay system, together with some indication of the relative efficiency? I recall that a load of approximately 200–300 kW is required by a terminal station such as Wenvoe for an output power of only 60 kW.

Is the requirement for standby supplies met at the repeater stations by two alternative Grid sources of supply, or by a standby generator? If standby generators are used, do the Post Office use the type of equipment which is at present on the market and in which a synchronous machine having a large flywheel is normally connected to the supply and is running continuously thereby under normal Grid supply conditions providing a measure of power-factor improvement? On disconnection of the normal supply the inertia of the flywheel allows the machine to operate as a generator until the standby Diesel plant is run up to speed.

The authors have referred to the automatic operation of the repeater stations, and I should be interested to have further information on the method of remote control and indication. Is control by d.c. impulses or by voice-frequency impulses? Are the pilots contained in a cable used solely for this purpose, or are they routed through normal Post Office exchange and repeater circuits? If the latter is the case, have the B.B.C. experienced any outages owing to interruption of the control-circuit pilots?

The authors indicate that if microwave working were also used for telephony circuits a spacing of 47 miles (the longest hop on the Manchester–Kirk o'Shotts route) would probably be too long to avoid fading. If we assume that a distance of 30 miles would be reasonably satisfactory for a microwave system, and we compare this with the repeater spacing of approximately 12 miles for normal multi-core cable carrier working, do the authors think that there would be considerable saving in capital with the adoption of the microwave system as against cable working?

Mr. H. J. R. Townsend (*at Cardiff*): To what extent does the quality of the picture suffer owing to its being "bounced" from point to point along the length of the country?

It is understood that, owing to the shortage of zinc, masts were aluminized and not galvanized. How is the aluminizing standing up to conditions of operation?

* DAWSON, G., HALL, L. L., HODGSON, K. G., MEERS, R. A., and MERRIMAN, J. H.: Paper No. 1623 R, January, 1954 (see 101, Part I, p. 93).

Mr. R. J. Harris (*at Portsmouth*): Could the authors give some idea of the life of a travelling-wave-valve amplifier? Could they also give some information on the desirability or necessity of using a sine-squared type of pulse for testing such links?

Mr. J. R. Hardwick (*at Portsmouth*): What limit of angular deflection of the axes of the 10 ft-diameter paraboloids, owing to the effect of windage on the structure of the supporting tower, can be accepted?

Mr. R. Goford (*at Portsmouth*): We can now presumably look forward to the increased use of s.f. links in the Post Office telecommunication system, and for example, now that we have the Rowridge link we might expect to be able to beam a transmission from there to centres of population in the vicinity, i.e. Portsmouth, Ryde, Portsdown, Southampton, etc., and filter out appropriate channels. The foresight of the Television Advisory Committee in 1948 had made that a possibility by catering for transmission of frequencies over a wide band.

I would like to pay tribute to the training facilities which have proved adequate to produce good maintenance staff from the rather raw material available. I think the projects and techniques involved, and the obvious efforts of the contractors and the Post Office to teach the men inspire them to show their talents to the best advantage, and I doubt whether we have ever been seriously disappointed at the staff's response despite the complexity and novelty of the work.

Interference problems in the Portsmouth telephone area have changed with the introduction of the Rowridge station, and I suppose this will be a common experience in any area where a poor field strength and signal/noise ratio have been experienced prior to the opening of a new station. Complaints of interference with television reception have fallen to about one-half of the previous number owing to the disappearance of fringe-area interference and the fact that existing users are now becoming very satisfied with the improved quality of reception. New users have not yet become critical, and they never should while present quality is maintained. Interference with medium-wave radio reception, on the other hand, has risen to about twice its previous value; prevalent sources are television i.f. and line-time-base interference. In cases where several television sets are having an effect on one radio set we have found it worth while to install screened aerial lead and generally improve the aerial and earth system of the radio set rather than to pay primary attention to the radiation from the television sets.

The Rowridge link has proved very reliable, and lost circuit time is very low indeed.

Mr. K. L. Rao (*India; communicated*): In Section 3.3 the authors state that, in order to minimize feeder losses, copper-weld waveguide is used for feeder runs. From the description of the system it appears that the feeder runs go up along the mast supporting the paraboloids at the top. No doubt the authors will be aware of the advances in technique made since the system was installed. Passive reflectors could be used with the paraboloids mounted at the base of the masts, and in many cases with advantage. A simple rearrangement may be necessary in the aerial and feeder arrangement.

In the London discussion Mr. Faulkner asked whether the authors have considered other forms of aerial system which might lead to reduced wind loads and thus reduced cost of towers. The use of passive reflectors not only reduces wind-loading problems but also enables paraboloids of greater gains to be constructed in addition to reducing losses due to waveguide runs to antennas and minimizing associated problems of the maintenance of aerial and feeder systems.

Messrs. G. Dawson, L. L. Hall, K. G. Hodgson, R. A. Meers and J. H. H. Merriman (*in reply*): The possibility of using a triode instead of the travelling-wave amplifier for amplification at 4000 Mc/s with a consequent reduction in the h.t. voltage required in the equipment was mentioned by one speaker. We understand that the gain-bandwidth product of the s.f. triode is such that three valves would be required to provide the 18 dB gain of a single travelling-wave amplifier, and that the circuit tuning procedure is more complicated.

We are convinced that the travelling-wave amplifier has definite advantages in terms of s.f. performance and ease of maintenance compared with the triode for this application.

A minimum average life of 3000 hours was specified for all microwave valves, and this has been exceeded in every case.

The video-frequency/group-delay characteristic of the overall system is shown in Fig. 14. Comparable results are achieved on coaxial-cable television links.

No pilots are used on the landline or radio link for the control of the overall system, and no outage time on the link has been caused by faults on the landline. The remote-control signalling system is described in the paper.

As stated by Mr. Warren, the average spacing of the s.f. repeaters is about 30 miles compared with about 12 miles for carrier cable and 6 miles for coaxial cable. It is not possible to give a general statement on the comparative capital costs of the different methods, as they differ considerably depending upon the specific requirements of a project and the terrain it traverses.

In reply to Mr. Townsend, the overall performance requirements are based on the use of four systems in tandem without introducing distortion in an amount which would be noticeable to the viewer using a good-quality domestic receiver. The performance figures achieved are given in the paper.

It is not possible to use the passive-reflector scheme suggested by Mr. Rao if only two radio frequencies are used for a 2-way system, as the attenuation between the two directions of transmission at a repeater station would be too small. We doubt whether the use of a reflector would reduce the wind-loading problems, since the reflector would need to be larger than the paraboloid for the same overall gain.

Automatic-starting arrangements for the standby engine generators are subject to maintenance routine check monthly and to date only one failure to start on a mains-supply failure has been noted. This occurrence was examined and found to be due to a wholly incidental and non-engineering error.

It is very likely that the continuously-running flywheel type of standby generator may be used on future systems.

In any discussion of the respective merits and demerits of staffed versus unattended operation and the related problem of extent and complexity of supervisory or control equipment, two factors must be borne in mind. The first is that in this system—the first of its kind in the world—a deliberate policy of over-generous supervisory-facility provision was adopted in order that possible operational problems, whose real magnitude could not be accurately gauged during planning, could be accommodated. The second is that, with a possible future tendency towards multi-channel operation on one route and the sharing between channels of supervisory services, the already favourable economic position of unattended operation is enhanced.

Overall electrical efficiency of microwave systems discussed by Mr. Warren is apt to be misleading. Station equipment loads upon the mains supply are of the order of 10 kW. The total radiated power is of the order of 2 watts. On the other hand, not only do the aerial systems used have a power gain of 40 dB, but it might also be argued that electrical power is accepted from the mains supply at the supply frequency with a nominal zero bandwidth and converted into radio-frequency power with a bandwidth of about 5 Mc/s. Our difficulty in conceding that overall efficiency is a significant factor may therefore be appreciated.

:1.317.79 : 681.142

Paper No. 1738 M
Oct. 1954

HIGH-SPEED ELECTRONIC-ANALOGUE COMPUTING TECHNIQUES

By D. M. MacKAY, B.Sc., Ph.D.

(*The paper was first received 29th April, and in revised form 9th July,* 1954. *It was published in October,* 1954, *and was read before the* MEASUREMENTS *and* RADIO SECTIONS *5th April,* 1955.)

SUMMARY

The paper describes circuits and techniques of measurement developed or use in an electronic-analogue differential analyser running at high epetition rates. The object has been to explore the upper limits to the peed conveniently attainable in various basic operations.

After a general description of the problem and the computer which as ultimately developed, these basic operations are treated in turn. recision methods have been developed for resetting and releasing ntegrators and altering the coefficients electronically at repetition ates up to 25 000 per second, making possible extremely rapid autoatic search processes of trial and error. A multiplier on the "quarter quares" principle has been developed, using biased diodes to control e feedback characteristic of an amplifier, giving an output which hows negligible phase-shifts at 50 kc/s.

Arbitrary functions are synthesized by approximation in diodeontrolled networks.

Essential to the efficient use of such high operating speeds has been e development of projective multidimensional displays with an dequate frequency response. Three- and 4-dimensional systems sing resistive resolvers are described, and the principles and techniques eveloped for accurate measurement at high speeds are discussed detail.

The final Section describes measurements made on each of the asic units to evaluate their performance.

It is concluded that the potential information capacity of electronicnalogue computing methods is much higher than can be realized by onventional techniques producing single solutions on a 2-dimensional isplay. With multi-dimensional displays and electronic programming, hundred- or even thousand-fold increase in information capacity is uite readily attainable with problems requiring systematic search rocesses for their solution.

LIST OF PRINCIPAL SYMBOLS

v_g = Instantaneous input-grid voltage of basic amplifier A.
v_o = Instantaneous output voltage of basic amplifier A.
a = Negative gain of basic amplifier A.
v_1 = Instantaneous input voltage to operating unit.
$\phi_{1, 2, 3}$ = Current/voltage functions of input elements P_1, P_2, P_3.
ψ = Voltage/current function of feedback element Q.
i_f = Feedback current.
$i_{1, 2, 3}$ = Current in input elements $P_{1, 2, 3}$.
V_X, V_Y = Control voltages operating diode switch.
R_g = Resistance between input grid and earth line.
R_f = Feedback resistance.
g_m = Transconductance of amplifier A.
V_w, V_x, V_y, V_z = Input voltages to 4-dimensional display (w is normally the dependent variable).
V'_w, V'_x, V'_y = Input voltages to 3-dimensional display (w' is normally the dependent variable).
V''_w, V''_x = Output voltages of 3-dimensional display (w'' is normally vertical).
V'''_z = Depth-co-ordinate output of 4-dimensional display.
V'''_y = Depth-co-ordinate output of 3-dimensional display.
α, β, γ = Rotation-angles in 4-dimensional display.
θ, ϕ = Angles of azimuth and elevation in 3-dimensional display.

(1) INTRODUCTION

Towards the end of the 1939–45 War, electronic-analogue computing techniques had reached an advanced stage of development for various military purposes. Their limited accuracy and restricted range of application, however, combined to support a growing opinion that the possibilities of such methods had been virtually exhausted and that they were destined to obsolescence by the advent of high-speed digital techniques.

The paper reports some of the results of a programme initiated in 1946 from a different aspect.* From a consideration of the current analogue techniques it was evident that great improvement in accuracy could not be expected, and that analogue differential analysers—like their prototype, the slide-rule—could hope to find applications only to those problems whose data were known or whose solutions were required to a comparable accuracy, say of the order of 1%.

But as the majority of physical problems belong to this range, at least in their exploratory stages, and as the theory of scientific information[1] (not to be confused with communication theory) suggested that an intelligent exchange of accuracy for speed was generally possible with a gain of informational capacity, it seemed not unprofitable to inquire how far the technique was, in fact, from the limits to performance set by physical necessity. Granted that accuracy could not be much improved, how far was it possible to increase speed without impairing accuracy to an unacceptable extent?

A high speed of operation would usually be of little value unless the operator were able to follow the activity of the instrument at speed. The goal of increased speed thus presented two main problems, the maintenance of adequate accuracy and the devising of adequate means of display. In the following Section the main elements of a computer developed as a result of this programme are briefly described. The remainder of the paper deals in more detail with some of the novel features embodied in its design.

(2) GENERAL DESCRIPTION

The computer is essentially a general-purpose differential analyser designed for cyclic operation at a variable solution rate with an upper limit of 25 000 solutions per second. It is controlled by a conventional crystal clock which provides four principal outputs:

(*a*) Main switching signals for cyclically releasing and resetting integrators and other units at the required rate, variable in binary steps between (25 000 ÷ 2^5) and 25 000 per second.

(*b*) Pulse-trains for calibrating the time-axis at intervals from 10 to 80 microsec.

(*c*) A binary group of subsidiary switching signals for automatically changing the coefficients of an equation in the interval between successive solutions.

(*d*) A master time-base generator.

Dr. MacKay is in the Wheatstone Physics Laboratory, King's College, University f London.

* Described more fully in the author's Ph.D. thesis, University of London, 1950, entitled: "The Application of Electronic Principles to the Solution of Differential Equations in Physics."

The computer is constructed on the unit principle. A set of identical general-purpose amplifiers provide a basis for all linear operations such as addition, integration and the like, being wired into circuit as each problem demands.

A set of general-purpose diode switching units driven by the main switching signals from the clock can be connected to any circuit which requires resetting to prescribed initial conditions before the start of each computing cycle.

Two coefficient-changing units controlled by the subsidiary switching signals from the clock can provide 32×32 successive combinations of any two chosen coefficients or parameters in an equation, so that 1 024 different trial solutions can be obtained before the system repeats itself.

A set of multiplying units capable of forming the instantaneous product of two rapidly-varying voltages make non-linear operations possible.

High-speed function generators driven by the clock provide the equivalent of an "input table" by means of which graphical data can be supplied at speed to the computer. The subsidiary switching signals already mentioned may be used to change the form of successive input-functions so generated, if required.

A multi-dimensional cathode-ray tube display capable of unambiguously depicting functions of two or three independent variables enables the order and form of successive solutions to be distinguished.

Specially designed measuring equipment, making use of null-observations wherever possible, is provided for use in setting up problems and for reading off information from the display.

(3) LINEAR OPERATING UNITS

(3.1) Amplifiers

The basic amplifier circuit (Fig. 1) comprises a single pentode, V_1, d.c. coupled to a cathode-follower, V_2, from which cathode-to-cathode feedback is derived through a preset control, RV_1. This enables the effective gain of the circuit to approach infinity under normal conditions. Frequency compensation in the network R_D, C_D, R_C, C_C keeps the gain substantially constant from zero to several hundred kilocycles per second.

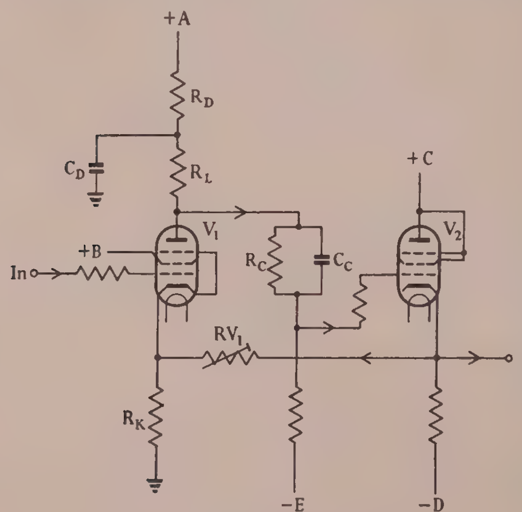


Fig. 1.—Basic amplifier circuit.

(3.2) Linear Operations

The well-known principle on which linear operations such as addition and integration are carried out is illustrated in Fig. 2.

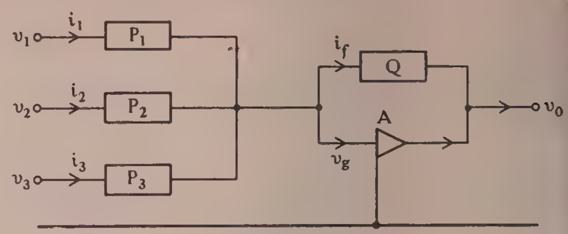


Fig. 2.—Typical operating unit with several inputs.

An amplifier A, of the type shown in Fig. 1, has an element Q connected between its input and output, and is linked to voltages v_1, v_2, v_3 by P_1, P_2, P_3. Since A is sign-reversing, it will tend to exert negative feedback via Q to maintain v_g at a small value. If the gain of A is $-a$, then $v_g = -v_o/a$. Now if the current i_1 in P_1 is a function $\phi_1(v_1 - v_g)$ of the voltage across it, and if the voltage $(v_g - v_o)$ across Q is a function $\psi(i_f)$ of the current through it, it may readily be seen that

$$v_g - v_o = \psi(i_1 + i_2 + i_3)$$
$$= \psi[\phi_1(v_1 - v_g) + \phi_2(v_2 - v_g) + \phi_3(v_3 - v_g)]$$

or

$$v_o = -[a/(a+1)]\psi[\phi_1(v_1 + v_o/a) + \phi_2(v_2 + v_o/a) + \phi_3(v_3 + v_o/a)] \quad . \quad (1)$$

When $a \gg 1$ and ψ and the ϕ's are linear operators, eqn. (1) becomes

$$v_o = -[\psi\phi_1(v_1) + \psi\phi_2(v_2) + \psi\phi_3(v_3)] \quad . \quad . \quad . \quad (2)$$

For example, if P_1 were an inductance L_1, P_2 a resistance R_2, P_3 a capacitance C_3, and Q a resistance R_f, eqn. (2) would become

$$v_o = -[(R_f/L_1)\int v_1 dt + (R_f/R_2)v_2 + R_f C_3 dV_3/dt] \quad . \quad (3)$$

Where P_1, P_2 and P_3 are resistances R_f/a_1, R_f/a_2 and R_f/a_3, eqn. (2) becomes

$$v_o = -[a_1 v_1 + a_2 v_2 + a_3 v_3] \quad . \quad . \quad . \quad . \quad (4)$$

To maintain accuracy at high frequencies it is often worth while in this case to shunt all resistances with small capacitances of proportionate reactance.

(3.3) Integration and Resetting

The most convenient integrator comprises a circuit of the type shown in Fig. 2 with P_1 a resistance R, and Q a capacitance C. Then (if P_2 and P_3 are omitted)

$$v_o \simeq -(1/RC)\int v_1 dt \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The principal problem is then to devise a method of discharging C at the end of each computing cycle and releasing the circuit with prescribed initial conditions at the beginning of the next cycle. To avoid loading the input circuit with the cathode-heater capacitance of diodes, the circuit shown in Fig. 3 was developed. D_1 and D_2 are normally held open by a positive voltage, V_Y, applied to R_2. When C is required to be discharged, the input V_Y is reversed in polarity, and a positive voltage, V_X, is applied to R_1. This voltage is large enough to bring the output v_o down to zero from its maximum possible positive value before the end of the resetting period. Thus even under the worst conditions there is a point at which D_1 and D_2 both conduct, and the input and output of A are thereafter held automatically at the same potential until the voltage V_Y is

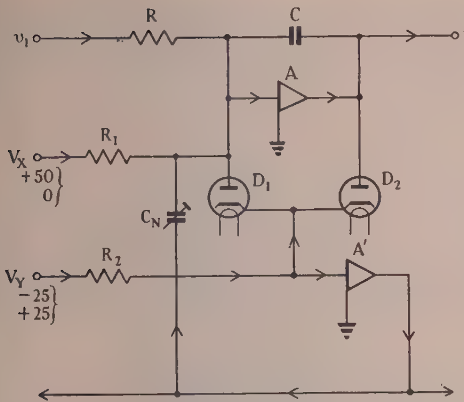


Fig. 3.—Integrator with diode-controlled resetting circuit.
A single neutralizing amplifier A' does duty for all integrators.

eversed in polarity, and V_X is once again reduced to zero, to tart a fresh cycle.

The self-capacitance of D_1 tends normally to transfer an ppreciable fraction of the switching voltage V_Y to the input of A. To obviate this the cathode voltage of D_1 is inverted in a ign-reversing amplifier A′ and fed back through a neutralizing apacitance C_N. A single amplifier A′ supplied with a typical nput serves to provide the neutralizing signal for all diode-witches. Its gain is $-1/4$, so that C_N can have four times the elf-capacitance of D_1 and hence be mechanically robust.

Diode switches are assembled in units complete with neutraliz-ng capacitors and other components, and are wired to the input nd output of standard amplifiers to convert them to integrators s required.

(3.4) Coefficient-Switching Units

The basic principle of the automatic coefficient switches is hown in Fig. 4A. S_{1-5} are diode switches of the type shown

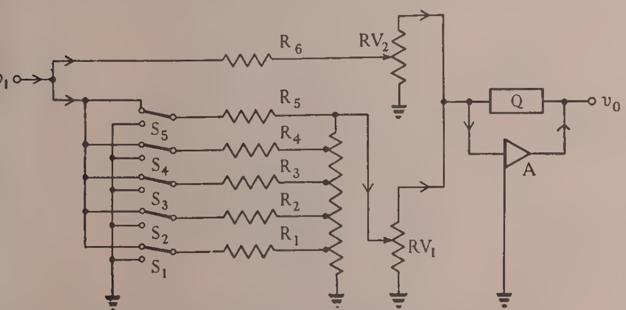


Fig. 4A.—Basic circuit of coefficient switching unit.

n Fig. 4B, and the network R_{1-5} is so designed that the proportion f V_1 passed to the input of the amplifier A when S_i is closed is roportional to 2^i. Thus the 2^5 possible combinations of S_{1-5} rovide 32 equal steps in gain. RV_1 makes it possible to adjust he magnitude of these steps, and RV_2 in conjunction with R_6 rovides a means of adjusting the minimum value of overall gain. The feedback element Q may be a resistance if only a variable-ain amplifier is required, or a capacitance (plus diode-switch) f an integrator is required with a swept time-constant of ntegration.

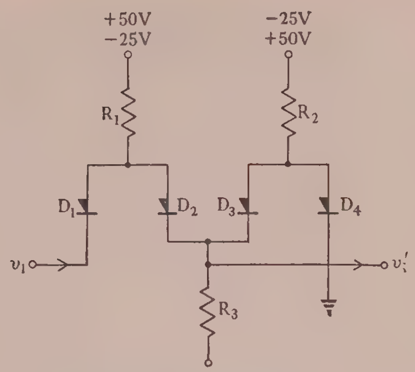


Fig. 4B.—Detail of typical diode switch shown in Fig. 4A.

RV_1 and RV_2 are connected in the positions shown rather than at the input or output terminals, to take advantage of the low impedance of the virtual earth at the input of A. This means that low-impedance potentiometers may be used to minimize phase-shift without the disadvantages of presenting a low input impedance to the preceding unit or of placing a heavy load on the output stage. It is easily shown[2] that the effect of a shunt impedance R_g at the input of A is to raise the output impedance only by an amount R_f/g_mR_g, where g_m is the transconductance of A.

The diode switch shown in Fig. 4B was devised in order to maintain a low output impedance under all conditions and to minimize the effects of self-capacitance and voltage drop in the diodes. Germanium rectifiers were used, driven by push-pull switching pulses alternating between +50 and −25 volts. Steady current flows through R_3. Under one condition, D_3 and D_4 are rendered non-conducting by a negative voltage applied to R_2, while D_1 and D_2 conduct and bring the output V_1' to the same potential as V_1. Under the other condition, D_1 and D_2 are rendered non-conducting by reversal of the voltage applied to R_1, while D_3 and D_4 conduct and bring the output to earth potential. Thus under all conditions the dynamic output impedance is only the forward resistance of two diodes in series, so that any unwanted signals transferred by stray capacitance are strongly attenuated.

At the same time, if the diodes are similar, the net voltage drop between the output and the operative input is zero.

(4) MULTIPLICATION

The high speed at which the computer was required to function demanded a higher frequency-response than could be provided by the electrodynamic multiplier (using crossed electric and magnetic fields in a cathode-ray tube) devised earlier.[3,4] The method finally adopted was based on the well-known principle of "quarter squares," whereby the product xy is obtained as the difference between $\frac{1}{4}(x + y)^2$ and $\frac{1}{4}(x - y)^2$.

To synthesize a square-law characteristic in an amplifier, the method shown in Fig. 5 was devised. A series of diodes, D_1–D_9, biased at different levels, control the effective transconductance of the feedback path of the sign-reversing amplifier A in such a manner that the gain varies linearly with the input current i_i.

The use of biased diodes to control the feedback path, rather than simply to synthesize a non-linear impedance in the input circuit,[5] has the advantage that a low impedance-level, and hence a short time-constant, can be maintained in the diode network without entailing a low input-impedance and a heavy (and variable) load on the preceding stage. It also makes it easy to compensate to a large extent for the self-capacitance of the

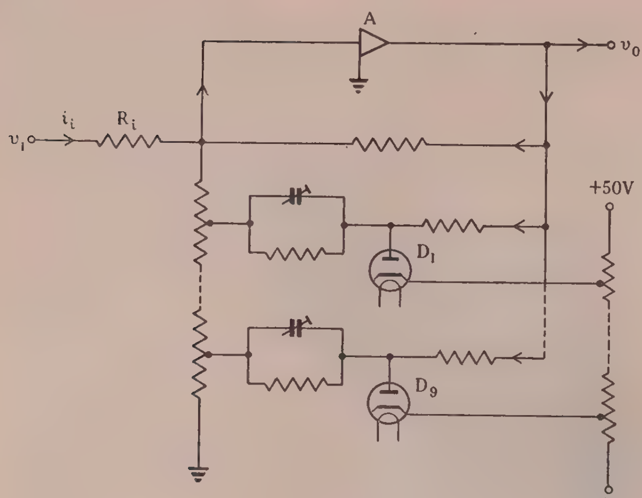


Fig. 5.—Basic circuit of square-law amplifier.

diodes by means of small trimmers across each of the feedback resistances.

No attempt was made to synthesize a symmetrical parabolic characteristic. The complete multiplier is shown schematically in Fig. 6, where it will be seen that the need for symmetrical characteristics is obviated by adding a constant input to each square-law amplifier and subtracting a suitable fraction of one input from the output.

The outputs of the non-linear amplifiers A_2 and A_3 are of the form

$$v_2 = -c_1(x - y + c_2)^2$$

and

$$v_3 = -c_1(x + y + c_2)^2$$

respectively, the sign of y being reversed in A_1.

The sign of v_3 is reversed in A_4, so that the input current to A_5 through R_{11} and R_{12} is proportional to $(4c_1xy + 4c_1c_2y)$. R_{13} is adjusted so that $R_{13}/R_{12} = 4c_1c_2$ and the term $4c_1c_2y$ in the input to A_5 from R_{11} and R_{12} is thus cancelled by the input from R_{13}. The total current in the feedback element of A_5 is therefore proportional only to $4c_1xy$. The constant c_2 is controlled by the voltage v_c supplied to R_7 and R_8. A_5 is constructed on the lines of a general-purpose amplifier, so that, for example the integral of the product xy can be obtained directly if required by the use of a capacitor with a diode switch unit as the feedback element Q instead of a resistor.

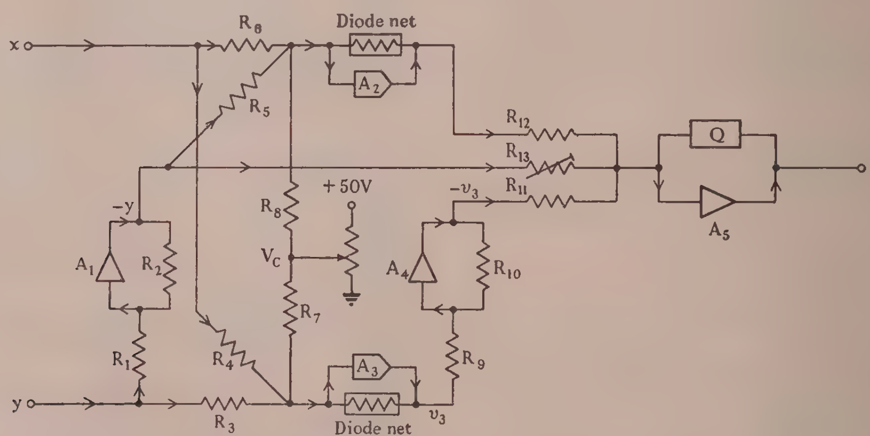


Fig. 6.—Block diagram of multiplier using two square-law amplifiers, A_2 and A_3.

(5) FUNCTION GENERATION

To generate functions with frequency components up to a few kilocycles per second, the photo-electric curve-follower[6,7] is adequate. For higher-speed work, however, a general-purpose function synthesizer was built, comprising a set of standard amplifiers and a set of germanium diodes and adjustable bias supplies.

By applying a standard sawtooth waveform to the set of biased diodes, discontinuities of voltage gradient can be obtained at any point along the time axis, and integrated as many times as required before combination in an adding circuit to form the required waveform. A simple example is shown in Figs. 7A and 7B, by which a "potential well," defined by

$$\begin{cases} y = -V \text{ (constant)} & \Big|_{x=0}^{x=a} \\ y = b(x-a)^2 - V & \Big|_{x=a}^{x=\infty} \end{cases}$$

is synthesized with the help of a single diode and integrator.

(6) DISPLAY

The projective principle[8,9] on which the 3-dimensional display is based is now well known, and the block diagram of the transformation unit is shown in Fig. 8. Potentiometer P_θ is a ganged sine-cosine unit which, via the adding amplifiers A_1, A_2, A_3 provides output voltages

$$V''_x = V'_x \cos\theta - V'_y \sin\theta$$

$$V''_y = -[V'_x \sin\theta + V'_y \cos\theta]$$

representing the projection of the vector (x', y') on a pair of axes rotated through an angle θ about the w' axis, where V'_x and V'_y are the voltages representing the two basal co-ordinates of the display and w' is the third variable. V''_x is applied directly

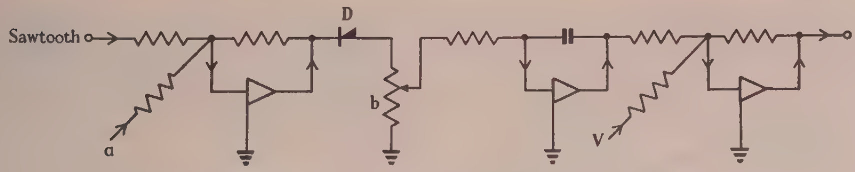


Fig. 7A.—Circuit used to generate the function shown in Fig. 7B.

) a push-pull d.c.-coupled amplifier feeding the X-plates of the ıthode-ray tube. V_y'' supplies one input to a second ganged ne-cosine potentiometer P_ϕ, which receives also the voltage $'_w$ representing the third co-ordinate.

P_ϕ in the same manner provides output voltages

$$V_w'' = -[V_w' \cos\phi + V_y'' \sin\phi]$$
$$V_y''' = -[V_w' \sin\phi - V_y'' \cos\phi]$$

epresenting the projection of the vector (w', y'') on axes rotated rough an angle ϕ about the x'' axis. V_w'' supplies a second .c. amplifier feeding the Y-plates of the cathode-ray tube. $_\theta$ and P_ϕ thus enable the display to be rotated in azimuth and evation respectively. V_y''' is proportional to the depth of the oint depicted by the co-ordinates (w'', x'') and is used optionally) modulate the grid and/or the final-anode voltage of the athode-ray tube, so as to add perspective by making the nearer ortions of the picture brighter and/or larger.

A simple extension of this principle was devised to enable -dimensional subjects to be displayed (Fig. 9). The projection om four dimensions to two is carried out in two stages, first om four to three, then from three to two as already described. lthough this involves more controls than the minimum mathe-atically necessary to resolve ambiguity in the display, it seemed e best way of aiding the intuitive perception of the resulting cture.

The vector represented by the four components (w, x, y and z) resolved in three steps by sine-cosine units, P_α, P_β and P_γ, entical with those already described. The first rotates the $, z)$ axes through an angle α about the plane (w, x) giving omponents y', z', represented by voltages V_y' and V_z'. V_y' is nt to the three-dimensional display unit, while V_z' and V_x e supplied to P_β. This rotates the (x, z') axes through an ngle β about the plane (y', w), producing a component V_x' hich goes to the 3-dimensional display, and an input V_z'' for P_γ. $_\gamma$ enables the (w, z'') axes to be rotated through an angle γ bout the plane (x', y''), producing the third input V_w' for the dimensional display, together with a voltage V_z''' representing e depth co-ordinate (in the fourth dimension) of the point presented in the 3-dimensional projection by (w', x', y'). This oltage V_z''' can conveniently be used for brightness modulation the display when V_y''' is used to modulate the final-anode oltage.

The 3-dimensional projection is displayed in the usual manner nd can be rotated by P_θ and P_ϕ to enable its "solid" form to be opreciated, and to resolve ambiguities. It is worth while in any such applications to use a stereoscopic pair of cathode-y tubes, adding small fractions of V_y''' (in opposite senses) to e x'' deflections of the two tubes.[8,9]

The usefulness of such displays depends naturally on the ensity of detail. The 4-dimensional display in particular is seful chiefly when some form of strobing can be used to eliminate om the picture all solutions except those that are of interest, g. those satisfying given boundary conditions. In many cases delay of one solution-period may be inevitable, but this is ften acceptable.

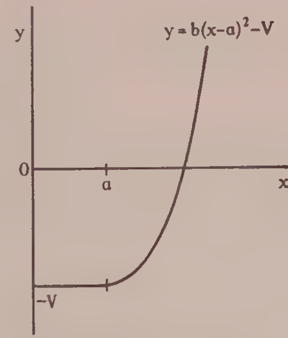


Fig. 7B.—Typical potential function generated by the circuit shown in Fig. 7A.

The parameters a, b and V are controlled by the voltages or potentiometer settings correspondingly marked in Fig. 7A.

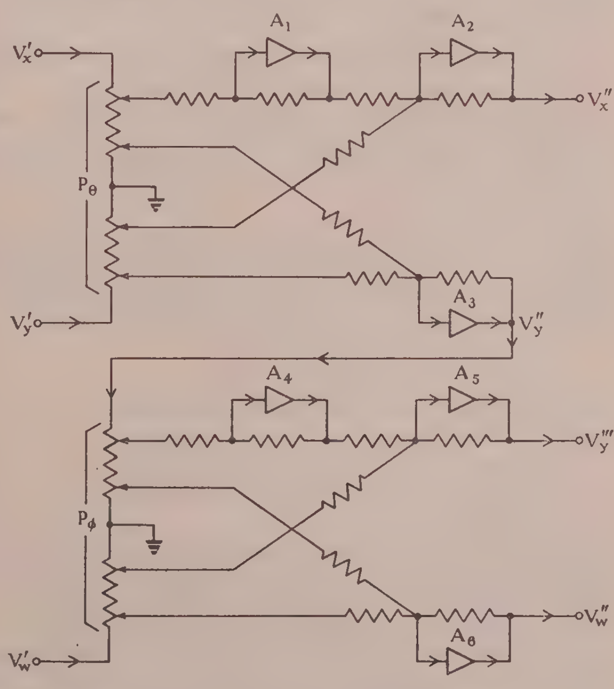


Fig. 8.—Basic circuit of 3-dimensional display.

As an accessory for use when null-observations on the main display can give sufficient accuracy, the transformation unit has provision for rapidly and repeatedly transferring all four inputs to four variable metered voltages.[8] The operation is performed by high-speed relays running at a sub-multiple of the solution rate, and results in the appearance in the display of a standard spot whose co-ordinates can be altered at will until the spot

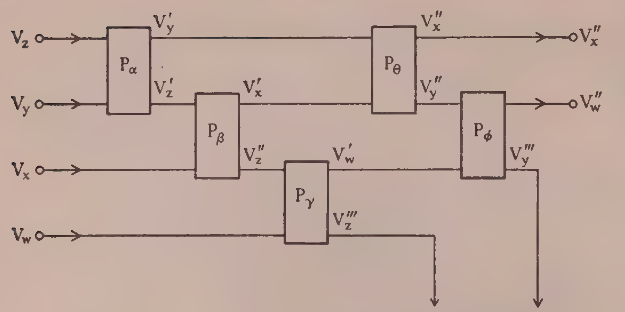


Fig. 9.—Block diagram of 4-dimensional display.
P_θ and P_ϕ represent the complete transformation units shown in Fig. 8.

coincides with the point whose position is to be measured. The rotation controls are used to reveal any errors due to parallax, and when these have been corrected the required co-ordinates can be read directly on the corresponding voltmeters.

(7) MEASURING EQUIPMENT

(7.1) General Principles

Since analogue computation is essentially the performance of an experiment, the principles and procedure adopted in the design of measuring equipment were those of physical experimentation. In particular it was decided that

(*a*) The number of independent standards should be kept to a minimum.
(*b*) Wherever possible, measurement should take the form of a comparison and null observation, enabling any distortion in the links following the comparison to be ignored.
(*c*) Setting-up, or at least final adjustment, of coefficients and parameters should be carried out by measurements on the complete stages *in situ*.

Measurement (and hence computing procedure) on these principles is not always as convenient as the reading of calibrated dials would be, but since the object was to explore the limits of techniques rather than to develop a desk computing instrument, questions of convenience were not much considered.

(7.2) Sources of Error

Sources of error in the operation of an analyser may be conveniently classified under three headings, namely

(*a*) Systematic errors in the setting up of equations.
(*b*) Systematic errors in the measurement of results.
(*c*) Random errors.

Under (*a*) are strictly included errors due to the fact that an element never precisely represents the operator for which it stands; this question has already been considered in Section 3. The chief errors which remain under (*a*) are those in setting up time-constants, coefficients and initial conditions when these are given.

Under (*b*) are included errors of observation of amplitudes, time co-ordinates, initial and final conditions when these are "unknowns" and coefficients adjusted by trial and error.

Under (*c*) come all errors which set a limit to the repeatability of results.

(7.3) Measuring Unit

In the analyser all measurements are referred to a single calibrated attenuator in a central measuring unit designed to minimize errors under these headings; the essential componen are shown in Fig. 10. The attenuator P has cathode-followi input and output coupling when required. A null-indicatir amplifier A_2 supplies a cathode-ray tube CRT_1, and a secon tube CRT_2 has both Y-plates available for direct comparisor when required.

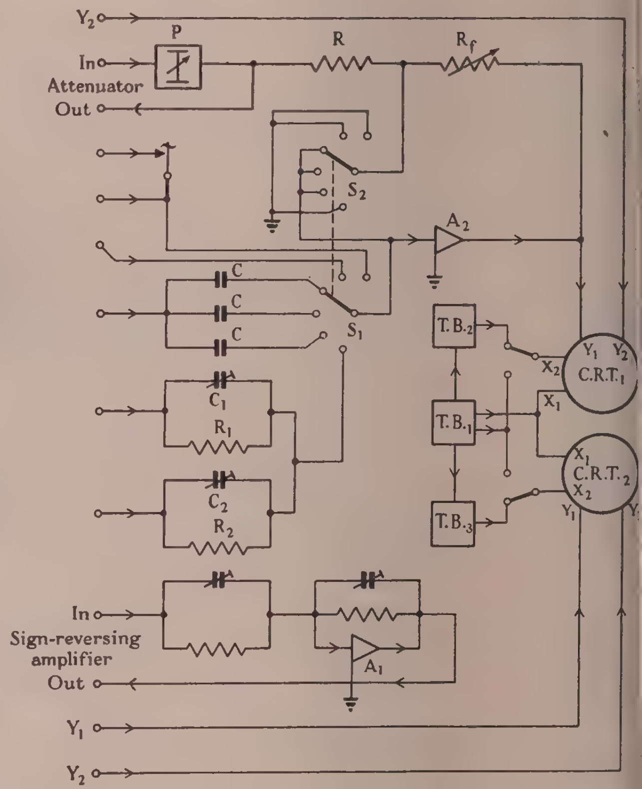


Fig. 10.—Block diagram of central measuring unit.

A universal time-base generator, TB_1, supplies both tubes, and two auxiliary generators, TB_2 and TB_3, can optionally be used to expand selected portions of the trace on either tube. A brightening-pulse amplifier enhances the brilliance of both tubes over the expanded portions of the trace and so identifies on one tube the portion shown expanded on the other.

(7.4) Measuring Techniques

(7.4.1) Gain.

The essential components used to measure the gain of a network are shown in Fig. 11. R_1 and R_2 are matched resistors (see Fig. 10) supplied from the input and output of the network N. The standard sign-reversing amplifier A_1 shown in Fig. 10 (gain $= -1$) is used to invert one signal if both are of the same sign. The attenuator P is introduced into the lead carrying the larger signal, and is adjusted until, at balance, its attenuation equals the gain (or reciprocal gain) of the network. Since the gain is measured *in situ*, it is possible by observing the computer display to verify that the attachment of the measuring apparatus has caused no serious disturbance. Cathode-follower probes were constructed to minimize such risks, but have seldom been found necessary. An obvious alternative is to set up the required conditions with the measuring apparatus connected, and then

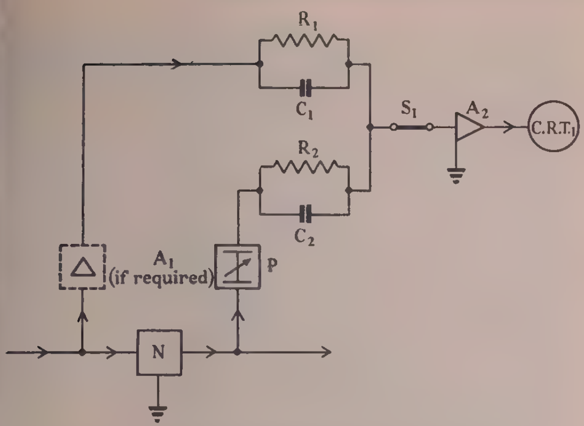


Fig. 11.—Method of measuring the gain of a network.

leave a simple equivalent dummy load in place of the latter when it is disconnected.

The input impedances used are all of 100 kilohms, to facilitate free interconnection of units, and small preset shunt capacitances of the order of 3 $\mu\mu$F are liberally used to maintain accuracy up to high frequencies. The balance of R_1C_1 and R_2C_2 is readily checked by interchanging their inputs, and the gain of A_1 can then be checked using R_1 and R_2 to add its input and output. A step input having a rise time of less than 0·5 microsec is used for such tests.

(7.4.2) **Voltage Gradients and Time-Constants.**

The basic standard of time in the computer is the pulse calibrator mentioned in Section 2, which delivers 10-, 20-, 40- and 80-microsec pulses. These may be used to determine integrating time-constants directly, by observation of the time taken by the output of the integrator concerned to become equal (and opposite) to a given steady input.

Where two integrators are used in series the product of their time-constants can similarly be found by coupling the input of the first to the output of the second through the standard sign-reversing amplifier A_1 and measuring the period of the resulting simple-harmonic oscillation, which is 2π times the square root of the required product.

For convenience, however, a substandard of time is incorporated in the measuring unit. This comprises a 100-kilohm high-stability resistance chain R in conjunction with one of three silvered-mica capacitors C (selected by the switch S_1, Fig. 10).

Rates of change of voltage are measured by applying the signal voltage to the free terminal of C and balancing it by applying to R, via P, a standard voltage step of adjustable magnitude and opposite sign produced in the crystal clock. At balance, the rate of change of the unknown voltage is $1/RC$ times the height of the calibrated voltage step.

In these switch positions a variable feedback resistor R_f effectively converts A_2 into a differentiating unit, in order to discriminate in favour of changes in voltage gradient, but its setting does not affect the null reading obtained.

To measure integrator time-constants it is necessary only to supply R and C with the input and output of the integrator in the running analyser (Fig. 12), adjusting the attenuator P until the potential of the junction of R and C remains steady. If the attenuator must be set to a ratio b for balance, the time-constant of the integrator is RC/b. If the time-constant is greater than RC, P must be used in the lead to C instead of to R.

This method has the advantage that the effective time-constant is measured at the operating frequencies, and that any error in integration is directly observed on CRT_1.

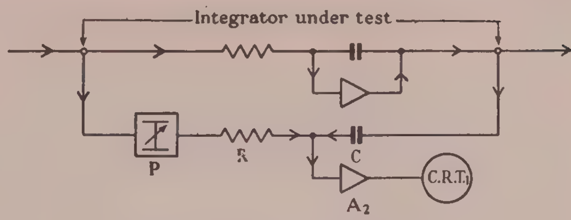


Fig. 12.—Method of measuring the time-constant of an integrator.

(7.4.3) **Amplitude.**

A similar null method enables ordinates of waveforms to be measured by comparison with the standard voltage step. In linear operations the absolute magnitude of this step need not be known, provided that initial conditions are measured in the same units.

Even in non-linear operations it is only the ratio between the ordinate concerned and the magnitude of the voltage standard in the non-linear element which is significant. The height of the standard step is accordingly adjusted to be a constant fraction of the voltage standard used for biasing the diodes in the multiplier, and is checked periodically against a step voltage generated from the bias standard by a vibrator, as shown in Fig. 13. The system is thus almost entirely scale-free.

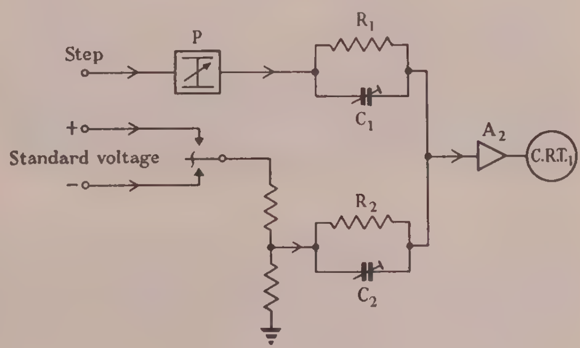


Fig. 13.—Method of standardizing a voltage step.

(7.4.4) **Summary of Measuring Technique.**

(7.4.4.1) *Setting Up.*

Irrespective of the accuracy of their component values, integrators can be set up to have specified time-constants within about 0·1% (the accuracy of the standards used in the central unit). Any systematic error (e.g. unwanted phase-shift) is detectable and can be measured or compensated for while the computer is running.

Numerical coefficients can be set up or measured with the same accuracy, and phase-shift can be detected or compensated for.

Initial rates of change of all outputs can be set up or measured in terms of the amplitude of a single standard voltage step.

(7.4.4.2) *Results.*

The same standard voltage step is used in the measurement of ordinates, and is itself set up against the reference voltage of the non-linear circuits, so that the system is scale-free.

Time co-ordinates are measured against crystal-controlled pulses, scale expansion being used to give more than adequate precision.

Unknown initial conditions are measured as in setting up. Final conditions are normally measured at a definite calibration point on the time scale rather than at the end of the solution period. They are most easily determined from the total effective amplitude of the inputs to an integrator rather than from the rate of change of output. The reference level of the input can be determined if the initial rate of change of output has been measured.

The accuracy of trial-and-error measurements is ensured by using the facilities described in Section 3.4 for adjusting the mean value and steadily reducing the range of variation of a trial coefficient until the required single value has been selected. This eliminates any errors introduced in the diode-switching system.

The close control of operation afforded by this technique makes the above errors small compared with those of such weak links as multipliers, whose errors cannot be easily compensated. In the present analyser the errors introduced in multiplication justify the measurement of many variables directly from the display, with a corresponding increase in the speed of operation.

Random errors due to supply-voltage ripple, valve noise and the like are in practice small compared with the systematic errors mentioned.

(8) TESTS OF PERFORMANCE

(8.1) Amplifiers

(8.1.1) Gain.

Without cathode-to-cathode feedback the amplifier shown in Fig. 1 was found to have a gain of 70 up to a half-value bandwidth of 750 kc/s, with a capacitive load of the order of 25 $\mu\mu$F. A by-pass capacitance of 0·001 μF across R_K raised the half-value bandwidth to some 1½ Mc/s, but led to instability when the amplifier was used in an integrating circuit.

(8.1.2) Phase-Shift.

Supplied with 100-kilohm input and feedback resistors by-passed by trimmer capacitors of the order of 4 $\mu\mu$F, the amplifier showed no perceptible phase-shift when supplied with a 60-kc/s square wave, after adjustment of the trimmers.

(8.1.3) Step Response.

Its response to a step function under the same conditions was tested by subtracting its output from its input in the (carefully matched) mixing network R_1, C_1 and R_2, C_2 shown in Fig. 10. The residual error-voltage, with a 0·002-μF capacitor in parallel with R_K, took the form of a spike of about 0·5% of the input in amplitude, with a duration of about 0·5 microsec.

The time-constant of response to a step function in the absence of feedback was observed to be of the order of 1 microsec.

(8.1.4) Stability.

The direct output-voltage level was observed to change by +0·2, −0·1, 0, 0, and −0·2 volt in response to increases of 10 volts in the A, B, C, D, E supply voltages respectively (Fig. 1), when 100-kilohm input and feedback resistors were used, the former being earthed.

(8.1.5) Output Impedance.

With a 100-kilohm feedback resistor and no input, the amplifier was found to have an output impedance of about 2 ohms, which is in good agreement with theory.

(8.1.6) Noise Level.

With 100-kilohm input and feedback resistors the noise level for a bandwidth of 100 kc/s was of the order of 2 mV. Vibration was found to give rise to voltages of the order of 5 mV, but all these random error-signals are, in practice, too small to be troublesome on the display.

(8.2) Coefficient-Switching Units

As already noted, the performance of the diode switches does not limit the accuracy of a final measurement, but tests were made to ensure that residual signals due to the switching pulses should not produce visible artefacts in the display.

With careful screening the output for zero input was reduced to a series of pulses of the order of 100 mV in amplitude and 5 microsec in duration, together with stepwise changes of the order of 5 mV. Since the coefficients are changed at the beginning of each resetting period (which is never less than 15 microsec in duration) these spurious effects are negligible.

(8.3) Integrators

The integrator circuit shown in Figs. 1 and 3 was tested for the effectiveness of compensation against break-through of switching signals and cathode-to-cathode feedback.

(8.3.1) Capacitive Compensation.

With the input connected to earth and trimmer capacitor C_N out of circuit, the output contained as an error term a voltage step 130 mV in amplitude, owing to the self-capacitance of D. With C_N in circuit the error term could be reduced to a transient of some 20 mV lasting only 1·3 microsec.

(8.3.2) Cathode-to-Cathode Feedback.

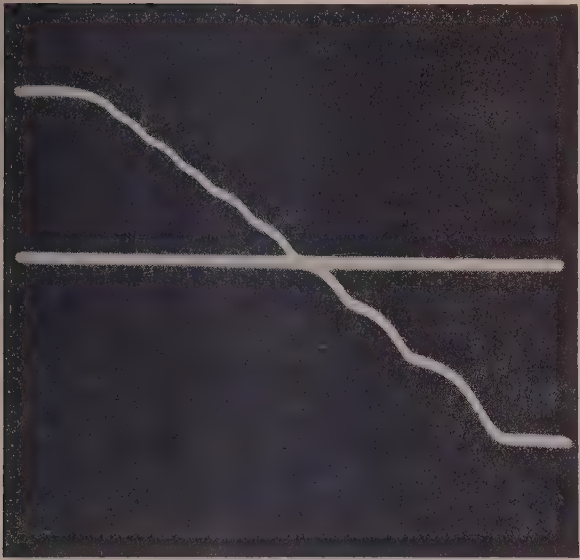
When cathode-to-cathode feedback had been optimally adjusted, it was found that the grid voltage varied by only some 5 mV in order to produce an output swing of 25 volts in each direction. With varying inputs a residual grid-voltage change is always present, in phase with the input. Its magnitude is determined by the integrating resistance R and the (irreducible) resistive component of the virtual input impedance of A under feedback conditions. For typical sinusoidal inputs between 5 and 100 kc/s it may vary between 5 and 30 mV peak-to-peak for an output of 15 volts, but since it is in phase with the input it introduces no error in the voltage across C, and its effects are not normally serious if the null methods described in Section 7.4.2 are used to set up the equivalent time-constant. The error term could be subtracted from the output either by inserting a suitable small resistance in series with C or by normal neutralizing methods, if the accuracy of other operations warranted this.

It should perhaps be said that, although optimal cathode-to-cathode feedback gives the amplifier A infinite (internal) gain, the presence of the heavy external feedback through C prevents any instability from arising from this cause, so long as C is regularly reset by the diode switch. It is possible to make the grid voltage (for a given output) reverse its sign, as R_f (Fig. 1) is reduced below the optimal value, without any sign of instability. The chief source of instability is in cumulative phase-shifts around the external loop at very high frequencies, particularly in integrators, which can give trouble in this respect even in the absence of cathode-to-cathode feedback.

(8.4) Multiplier

(8.4.1) Square-Law Circuit.

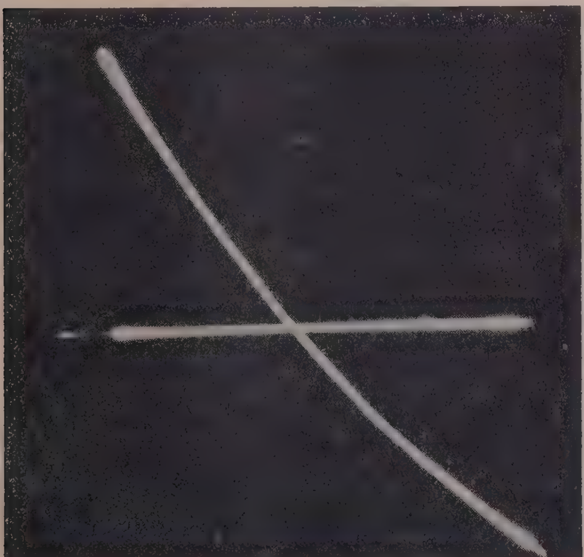
The accuracy of a typical square-law circuit may be gauged from Fig. 14, which shows the gain as a function of input voltage. The graph was obtained by supplying a linearly rising input and



ig. 14.—Oscillogram of gain versus input voltage for the square-law amplifier shown in Fig. 5.

lifferentiating the output, and it indicates an accuracy in slope f the order of $\pm 2\frac{1}{2}\%$. Higher accuracy could evidently be btained by using more diodes or by reducing the voltage scale o as to profit more from the curvature of the diode characteristic. An alternative device for which provision was made, but which as not as yet been tried, would be to inject into the bias circuit high-frequency sawtooth signal just large enough to make the ransition zones of neighbouring diodes overlap, and high enough n frequency to make the transitions practically continuous on he normal time-scale.

An indication of the frequency response is provided by Fig. 15, which shows the input/output characteristic drawn on a cathode-ay tube at 50 kc/s. The fact that the trace is only slightly thickened indicates the effectiveness of the phase-compensating trimmers shown in Fig. 5.

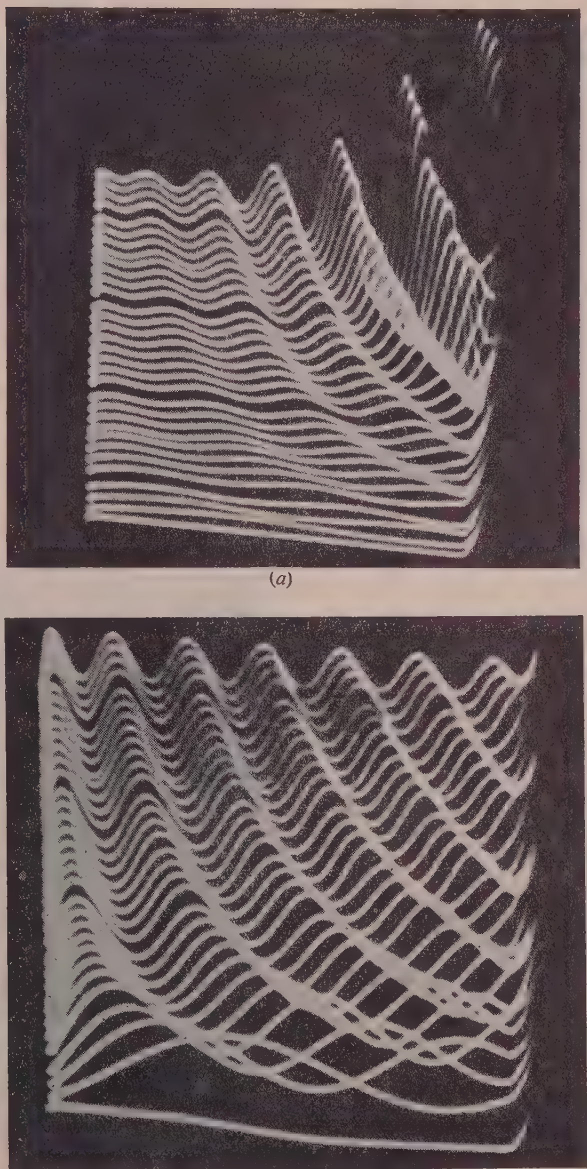


ig. 15.—Input/output characteristic of square-law amplifier at 50 kc/s.

(8.4.2) **Overall Characteristics.**

The complete multiplier was tested by measuring the gain between one input terminal and the output, as a function of the (steady) voltage applied to the other input terminal. For steady inputs between ± 5 volts, the deviations from linearity of the gain/input characteristics were of the order of $\pm 1\%$ of maximum gain. Over a range of input between ± 10 volts, the maximum error found was of the order of $\pm 2\%$ in each case.

With one input earthed and the other supplied with ± 10 volts a.c., a zero-error signal of the order of $\pm 1\%$ of full scale was



(*a*)

(*b*)

Fig. 16.—Three-dimensional display of a family of s.h.m. solutions.

(*a*) Showing exponential increase in amplitude due to phase shift at high frequencies. (Total time-base = 40 microsec.)

(*b*) Showing effects of phase compensation.

observed at low frequencies, rising to $\pm 5\%$ of full scale at 50 kc/s.

With a steady input of 10 volts to one terminal and ± 10 volts a.c. at 50 kc/s to the other, the input/output characteristic showed negligible phase shift.

(8.5) Computing Networks

Two integrators and a coefficient-switching unit were set up to solve the s.h.m. equation

$$d^2y/dt^2 + \omega^2 y = 0$$

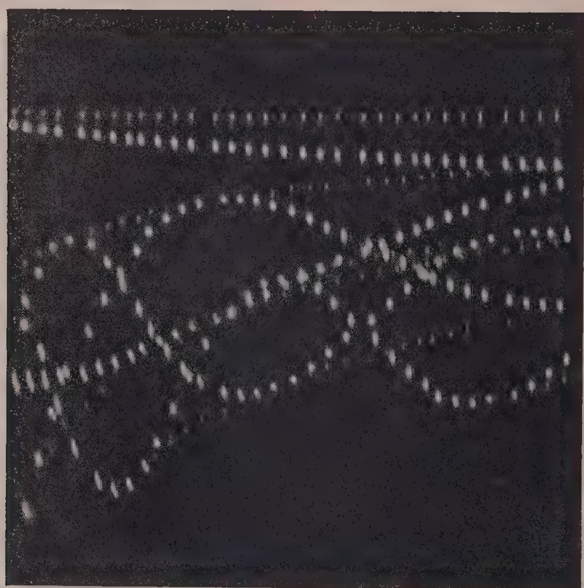
in order to test the performance of a simple network of units at high frequencies. It was found as predicted by theory[4] that the limited bandwidth of the sign-reversing amplifier in the switching unit caused the amplitude of higher-frequency solutions to increase exponentially with time. This is illustrated in Fig. 16(*a*), which shows the 3-dimensional display of 32 solutions for a succession of values of ω^2. It seemed likely, however, that to a first order the effect could be regarded as equivalent to that of a capacitance in parallel with the feedback resistance of the sign-reversing amplifier, and it seemed worth while to try a simple neutralizing scheme. The output of the amplifier was inverted in a further unit with a gain of $-1/4$, and fed back to the input grid through a small trimming capacitance.

An impression of the result may be gained from Fig. 16(*b*).

Solutions which normally [Fig. 16(*a*)] would build up to

(*a*)

(*b*)



(*c*)

Fig. 17.—Solutions of s.h.m. equation for varying values of ω^2.

(*a*) Repetition rate of 1 562·5 solutions per second.
(*b*) Repetition rate of 12 500 solutions per second; calibration spots are 10 microsec apart.
(*c*) Side view of solutions shown in (*b*).

saturation in a few cycles can now be held constant in amplitude within 1% at all frequencies up to 120 kc/s for a single setting of the neutralizing trimmer.

Since the s.h.m. equation sets the most stringent test in this respect, serious error from this cause would seem to be avoidable up to about 200 kc/s. Above 200 kc/s the compensation was not satisfactory, owing no doubt to the many other causes of phase-shift significantly operative at these frequencies.

(8.6) Displays

Figs. 16(*a*) and 16(*b*) were taken with a recording camera from a small cathode-ray tube. A better indication of the performance of the display is given by Figs. 17 and 18, which

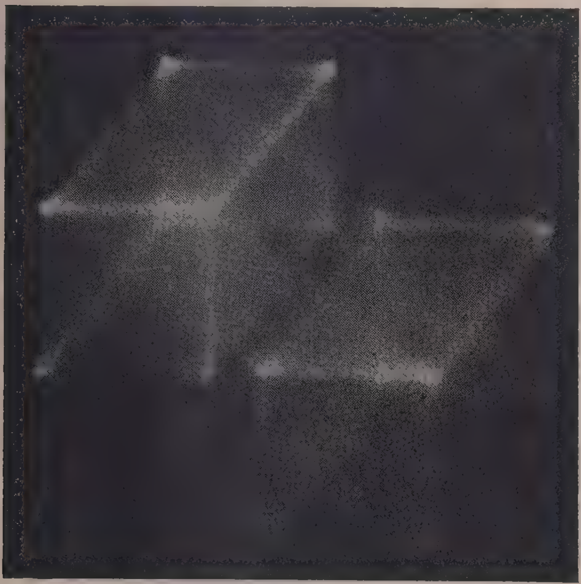


Fig. 18.—Tesseract (4-dimensional "cube") on 4-dimensional display.

have been taken from the main 12in display tube. Fig. 17 shows solutions of the s.h.m. equation for various values of ω^2 at repetition rates of 1 562·5 per second and 12 500 per second respectively. In the first the calibrator pulses accentuate every second and fourth 10-microsec mark. In the second, identical pulses mark each interval of 10 microsec. Fig. 17(*c*) shows a side view of the surface shown in Fig. 17(*b*). The calibration spots here can be seen to trace out isochronous curves of amplitude as a function of the third variable ω^2—a typical example of the new kind of facility offered by a multi-dimensional display. It will be seen that solutions are grouped in fours and eights on the display for ease of identification.

Pictures on the 4-dimensional display are not readily understood unless the rotation controls can be used to resolve ambiguities, but Fig. 18 shows one of the simplest subjects, the tesseract, or 4-dimensional analogue of the cube. From the photograph it is easily seen that from any point there are four basic directions in which to move. That these are orthogonal can be shown only by rotation of the subject; but if the fact is granted, Fig. 18 can be seen to illustrate other interesting features of 4-space, such as the existence of orthogonal pairs of planes which meet in only one point. The Figure was produced simply by supplying the four inputs of the display with four asychronous oscillatory signals.

(9) CONCLUSIONS

The principal conclusion drawn from the work described is that a large reserve of information capacity has hitherto* been untapped by conventional electronic differential analysers. With the novel high-speed techniques described, something nearer to the theoretical limit for normal electronic components can be achieved. In linear problems involving only constant coefficients or parametric variables, it has been found that solutions containing frequency components up to 200 kc/s can be handled by suitably compensated operating units without undue difficulty.

As to accuracy, compensated amplifiers have been found to introduce less than 0·1% distortion at frequencies up to 1 Mc/s or so. The simple integrators used have proved capable of an accuracy better than 0·2% at all frequencies up to 100 kc/s. The multiplier constructed had an accuracy of the order of $\pm 2\%$ and introduces negligible phase-shift at frequencies up to 50 kc/s. The effect of finite bandwidth in a complete system can be neutralized for frequencies up to 120 kc/s to within 1% in the case of the s.h.m. equation.

Accordingly, it is concluded that, for linear problems involving only parametric variables, a solution rate of $12\frac{1}{2}$–25 kc/s may often be practicable. High-speed switching techniques developed for these rates have proved satisfactory, enabling the values of parameters to be altered and initial conditions reset within 15 microsec, up to 25 000 times per second. At rates below 10 000 per second an overall accuracy of the order of $\pm\frac{1}{2}\%$ or better is attainable in linear problems and $\pm 2\frac{1}{2}\%$ in problems requiring the multiplication of variables. The evidence suggests that the latter figure could be considerably improved by the more generous use of diodes in the non-linear circuits.†

The one hundred- to one thousand-fold increase in speed over conventional analysers attainable in this instrument has been made worth while by using multi-dimensional displays to make an adequate volume of information space accessible to the operator. The extension of the projective principle to the display of 4-dimensional data provides a new facility of which the possibilities are largely unexplored, but which has obvious applications to complex trial-and-error problems.

Since improved performance has here been achieved more through the use of novel circuit techniques than by meticulous care in design, it is not felt that the limit of performance has been closely approached by the present equipment, and the performance figures quoted should be considered as stimulants rather than as standards. Probably not more than 10% of available information capacity is utilized by the apparatus in its present form.

It appears in fact that analogue technique has more to offer. The programme here described has done little more than open up the possibilities.

(10) ACKNOWLEDGMENTS

The author is much indebted to a number of past and present members of the technical staff of King's College for their assistance in construction of the apparatus described.

(11) REFERENCES

(1) MacKay, D. M.: "Quantal Aspects of Scientific Information," *Philosophical Magazine*, 1950, **41**, p. 289.

(2) MacKay, D. M.: "Application of Electronic Principles to the Solution of Differential Equations in Physics," Ph.D. Thesis, 1950, University of London, Appendix 2.

* At least in 1950, when the original thesis was written.

† Work on the computer was unavoidably shelved in 1950, but has recently been resumed by others in this Laboratory; and the outcome of such improvement will be reported in due course.

(3) Deeley, E. M., and MacKay, D. M.: "Multiplication and Division by Electronic-Analogue Methods," *Nature*, 1949, **163**, p. 650.
(4) Macnee, A. B.: "An Electronic Differential Analyser," *Proceedings of the Institute of Radio Engineers*, 1949, **37**, p. 1315.
(5) Chance, B., *et al.*: "A Quarter-Square Multiplier using a Segmented Parabolic Characteristic," *Review of Scientific Instruments*, 1951, **22**, p. 683.
(6) MacKay, D. M.: "A High-speed Electronic Function-generator," *Nature*, 1947, **159**, p. 406.
(7) Mynall, D. J.: "Electrical Analogue Computing," *Electronic Engineering*, 1947, **19**, p. 178.
(8) MacKay, D. M.: "Projective Three-dimensional Displays," *ibid.*, 1949, pp. 248 and 281.
(9) Parker, E., and Wallis, P. R.: "Three-dimensional Cathode-Ray-Tube Displays," *Journal I.E.E.*, 1948, **95**, Part III, p. 371.

DISCUSSION BEFORE THE MEASUREMENTS AND RADIO SECTIONS, 5TH APRIL, 1955

Mr. D. J. Mynall: Function generation receives relatively little attention in the paper, and the method described as a higher-speed alternative to the photo-electric curve-follower cannot always be used as a substitute. When a network contains integrators, it may be arranged to generate a pre-assigned function of the variable identified with time, but it cannot do this for any other variable in the system. The photo-electric curve-follower does not have this limitation. Diode networks not containing integrators may be employed, of course, but it is then usually necessary to use a considerable number of diodes in order to approximate the desired function sufficiently closely. Further complications arise when the function contains maxima or minima. The paper may give the impression that function generation is simpler than it often turns out to be in fact.

For basic amplifiers, I found it better to start with a relatively high gain (100000, as compared with the 70 quoted in the paper), even when increase of the gain by positive feedback was employed. With an initial basic gain of 100000 it is still desirable to use positive feedback to obtain the best performance from integrators, but one has the advantage that drifting of circuit conditions has far less effect on performance than when the bulk of the necessary gain is obtained by positive feedback. Furthermore, a negative-feedback sign-reversing amplifier using a basic amplifier of high gain has, *ipso facto*, a wider frequency pass-band than one using a basic amplifier of low gain, and effects such as that shown in Fig. 16(*a*) are usually of negligible proportions, thus requiring no special adjustment of the amplifier such as the phase compensation described in the paper.

A point worth noting with this type of computer is that it is extremely easy to find out what is happening in all parts of the circuit merely by connecting various points in turn to the display. This means that precise computation of all scale factors prior to set-up is unnecessary, since one can rapidly find out whether any element is over- or under-loaded and correct accordingly.

There is no fundamental high-accuracy requirement with three-dimensional display (the chief function of which is to assist one to understand the significance of what is on the screen), and an entirely satisfactory practical simplification of the apparatus can be used when limitation of the angles of rotation is acceptable. This is effected by using linear potentiometers, in conjunction with sign-reversing amplifiers, appropriately connected to give approximate sines and cosines. The need for special sine-cosine potentiometers is thus avoided.

Four-dimensional display is fascinating, but there is a considerable risk of including enough information on the display to give oneself mental indigestion. A possible field of use for 4-dimensional display would be as an educational aid in the teaching of mathematics.

Mr. E. H. Cooke-Yarborough: I am very glad that this work has been done, because I have thought for a long time that, in an ordinary analogue computer working in real time, valves working in circuits with time-constants of many seconds were not really earning their living. Indeed, the valves themselves often seemed to express their dissatisfaction with the situation by oscillating at a frequency of several megacycles per second.

I thought at first that the author had also overcome one of the biggest difficulties in analogue computers—d.c. drift—because his computer goes through the whole process so quickly. However, it appears that d.c. drifts will not only displace the display bodily but will also change the shape of the picture by altering the multiplying factors. It therefore seems that the circuits must not drift appreciably in the time taken to set them up, interpret the picture and make the necessary measurements. Probably this computer is therefore not much less sensitive to d.c. drifts than an ordinary analogue computer working in real time.

In this respect I note that Fig. 1 shows no provision for eliminating spurious direct voltages produced by variations in valves or component tolerances. If valves are selected at random and inserted into the circuit and a feedback resistor is connected from the output cathode back to the input grid, then, with no input signal, the input and output voltages will be equal but not necessarily zero. If the output is not zero for zero input signal, difficulties will arise in the coefficient switching unit shown in Fig. 4(*a*), in which resistors are switched between the amplifier output point and earth. I presume, therefore, that the author has used some expedients which he has not mentioned to bring the output voltage of his amplifiers to zero for zero input signal.

From Section 8.2 it would appear that the germanium-diode switches introduce errors of the order of 5mV. This presumably represents the difference between the forward voltages of the two diodes in the switch. I am surprised that the performance is as good as this, for we have noticed differences of as much as $\frac{1}{2}$ volt between diodes. Does the author select his diodes to achieve these good results?

If diode switches are as good as this, the author might be able to take this switching technique further. One of the things that make many analogue computers such large complex devices is the need to provide d.c. compensation in the amplifiers, which is usually done by use of electro-mechanical chopper amplifiers. Could not the germanium-diode switches be used in a similar way, probably providing a d.c. stability intermediate between that of an uncompensated d.c. amplifier and that of one compensated by the conventional chopper amplifier?

Regarding the multiplier, I am a little dubious about adding a large voltage to the input to make sure the resultant voltages never go negative, and then subtracting a large voltage afterwards. When small voltages are being handled by the multiplier, it appears that the obtaining of an accurate result depends upon an accurate measurement of a small difference between large voltages, and this may detract from the accuracy.

Incidentally, there appears to be considerable difficulty in defining the accuracy of an analogue computer. The user is apt to think that the figure quoted is a guaranteed accuracy and that results can be expected to lie within the limits quoted. In practice, however, errors can easily become large, particularly if

ome quantity is small or becomes small in some part of the omputation. It is safer, therefore, to regard the accuracy gures which may be quoted as representing the lower limit of ne likely errors rather than an upper limit.

Lt.-Cmdr. F. R. J. Spearman: Both this paper and another* iven recently before the Royal Aeronautical Society confirm ny belief that analogue computers have a big future in the omputing field.

I have been mainly concerned with the slower real-time single-hot analogue computers, such as are used in aeronautical esearch, in which real time is used because real components an then be used in the system. One point which impresses me the great difference which there can be in the design of analogue-omputer components, even amongst those which might be con-dered to be standard. The positive-feedback amplifier described the paper is very different from the high-gain drift-corrected .c. amplifier of the type used in real-time analogue computing.

It is this difference in the components that makes the develop-nent of a general-purpose computer very difficult. There is a eal requirement for a general-purpose computer, but it is very ifficult to specify. Does the author consider his positive-feed-ack amplifier suitable for use in such computers, or would the rift be a major problem? Would it be possible to fit some type f drift corrector to his type of amplifier without either com-licating it unduly or causing any oscillation?

A start is now being made in this country in producing com-ercially some items of analogue-computing equipment which ill meet the limited objective of simulating, either in real time r repetitively, events such as occur during the high-speed flight f aircraft or guided missiles, either controlled automatically or y human operator. The design of an amplifier with low drift nd high gain over a wide bandwidth is one which would aterially assist the design of a general-purpose computer.

The author's efforts appear to have been devoted mainly to eveloping techniques for high-speed analogue computing, for hich he has contrived some very ingenious methods and circuits nd also the means of presenting the results. In all these matters e presentation of results is very important: one can use an nalogue computer to produce so many results that there is no me in which to analyse them, so there must be some means of resenting the results. However, to persuade others to use such quipment, it is necessary to show that the techniques developed n be usefully applied to a number of different problems. I do ot see this in the paper. The author has since given some amples on how to apply his computers to different types of roblem, and I should be grateful for further details of problems which he thinks they could be applied.

Lastly, has the author attempted to assess whether such a omputer has sufficient uses to justify its design as a commercial strument for use either in industrial research or in any labora-ry which provides computing facilities?

Mr. E. G. C. Burt: For the "quarter squares" multiplier the thor uses a pedestal voltage V_c to ensure that the input to each uaring unit is always positive, whatever the size of the inputs and y. This involves subtracting a voltage proportional to y om the output, and it also assumes that the two circuits are fficiently balanced to eliminate the square of the pedestal ltage which arises from each. Since V_c must be greater than e arithmetical sum of the maximum values of x and y, the nge of voltage input to the squaring circuits is unnecessarily eat. I am not convinced that anything is gained by this ocedure, since, although only two diode networks are used, e circuit requires five amplifiers.

In an alternative arrangement which we use a good deal at the Royal Aircraft Establishment, two symmetrical diode charac-teristics are provided, i.e. four diode units, but only three ampli-fiers are required. This gives the correctly signed product xy when x and y are unrestricted in sign. Either the positive or the negative product is selected by a switch, so that the multiplier includes its own sign-reversing element—a useful facility in an analogue computer. Either output is available at a low impedance.

We use drift-corrected d.c. amplifiers with a large voltage swing (± 50 volts), and accuracies of the order of $0 \cdot 2\%$ have been achieved.

As the author points out, it is possible to obtain a time-dependent function of the product, such as the time integral, rather than the product itself, by suitably modifying the feedback element of the final amplifier.

It seems to me that the non-linear function generator is useful only when the independent variable x is directly related to time. Non-linear systems involve non-linear functions of the dependent variables; the diode technique, as used in the multiplier, can be extended to include such functions even when they are not monotonic. In fact, at the R.A.E. we have used, and are using, such units to generate various kinds of functions; for example, to provide trigonometric functions for axis transforma-tion, and to simulate non-linearities arising in aerodynamic equations, where accuracy and speed of response are of great importance.

Mr. K. W. Thwaites: The author has succeeded in showing that the development of analogue computing techniques is far from complete and that further work in this field will be amply repaid. The advantages peculiar to the analogue system are well known, particularly the ease with which problems may be set up without the need for long experience of programming, and, in certain applications, the speed of solution, which may con-siderably exceed that of an equivalent digital computer.

The significant feature of the author's work is his success in resetting integrators and the values of coefficients in the computer chain at high speed and at predetermined intervals. Could not this technique be carried further if, instead of changing coefficients in a predetermined order, the choice of coefficients could be determined by certain characteristics of the answer obtained from the previous computation? A very powerful weapon would then be available for the rapid solution of problems by the process of iteration. The fitting of mathematical functions to practical curves is an example of the type of problem which could be handled.

I note that a positive feedback amplifier has been used as the fundamental unit of the computer, and previous speakers have already asked whether this is good enough in terms of d.c. stability. The paper describes this amplifier applied to adding circuits and integrating circuits, but not to a differentiating circuit. The latter case provides a more difficult condition for stability, since the phase margin is reduced in the presence of capacitive input and resistive feedback. Has the author achieved stability in this type of circuit while still maintaining a useful forward gain?

Does hysteresis of the condensers give rise to any trouble in the resetting cycle of the integrators, since the interval between resetting and integration must, in this application, be very short?

Mr. C. A. A. Wass: The Royal Aircraft Establishment has a collection of analogue computers claimed to be the largest in Western Europe, all of which work substantially in one-to-one time. This is not because we are unaware of the attraction of doing things rapidly, but there are good reasons. For instance, one can include parts of a real system; we are interested in aeroplanes, and it is difficult to describe in mathematical terms the behaviour of the servo motor driving the rudder. We can remove the motor, and put it in the machine, and then subject

* DIPROSE, K. V.: "Analogue Computing applied to Aeronautical Problems," *rnal of the Royal Aeronautical Society*, 1955, 59 (to be published).

it to input signals of a much more realistic kind than can be produced on the bench.

There is another important aspect of one-to-one time scales: with all kinds of repetitive simulators—the kind which give a steady picture on a screen—the experiment is limited to rather brief stimuli. The author has used step functions throughout, and this is adequate for his purpose, but for many other purposes a step function is rather inappropriate. For example, we are interested in calculations about aeroplanes in flight; aeroplanes do many peculiar things, but they seldom do step functions. The manoeuvres are very often complicated functions of time, which become very difficult to simulate on an accelerated time scale.

By using repetitive devices, there is a gain in apparent convenience: one can turn a knob and watch things happening rapidly. For our kind of problem, however, this is not necessarily an advantage. We are interested in the kind of problem for which one-to-one time-scale computers can produce answers faster than digital machines by a ratio of something like 500 or possible 1000.

I have one small point concerning the use of diodes. The author says that he devised an arrangement with biased diodes. I do not think he intended to suggest that he invented the diode technique, for the idea was comparatively well known before the date of any of his references. My attention was recently drawn to an American patent, according to which the germ of the idea of using a diode or even triode with bias on it to obtain various shapes was handled by the United States Patent Office in 1919.

Professor C. Holt-Smith: In generating our solutions at a very rapid rate we are solving one problem but creating another and perhaps more difficult problem of how to handle the data as they become available. We must be able to make use of the information at the rate at which it is generated, otherwise we are creating difficulties inside the computer for no profitable purpose.

If we generate information at the rate of 12500 times per second and generate perhaps 30 solutions on the display, we are generating a whole class of solutions in the order of a millisecond. If we are then to superimpose a thousand of them before we take action, we might equally have generated it once in a second, made the design problem very much easier, and taken a photograph.

In order to justify the difficulties of making an analogue computer operate in 40 microsec, we must have a way of efficiently utilizing thousands of solutions per second, and this seems to be a very real and important problem that must be studied before we can take full advantage of the material in the paper.

Instr. Cmdr. D. K. McCleery: In Section 3.3 the author says that the diodes D_1 and D_2 (Fig. 3) are "normally held open by a positive voltage applied to R_2." I understood that when a valve was "open" it was passing current: a positive voltage applied to R_2 shuts down the valves (cathodes positive), according to my way of thinking. In these pulse circuits "everything opens and shuts," so I think it is important to avoid any ambiguity. It seems clear that the author is thinking of the analogy of a mechanical switch, not a valve, and I agree that an "open" switch should not pass current. The word "valve" is, of course, another analogy from mechanics, but the snag is that it works the other way, whether gases, fluids or electrons are passing through it. I suggest that the terms "off" and "on" would resolve this ambiguity, and would apply equally to a mechanical or an electronic switch.

THE AUTHORS' REPLY TO THE ABOVE DISCUSSION

Dr. D. M. MacKay (*in reply*): Mr. Mynall and Mr. Burt are right in emphasizing that integrators cannot be used in generating functions of variables other than time. But, as suggested in the original thesis (Reference 2), the same general-purpose unit can be used for stepwise approximation in such cases, and maxima and minima can be produced by subtracting part of the input from the output. The curvature of the diode characteristic fortunately reduces the number of elements required in making such approximations to smooth functions.

I imagine that Mr. Mynall's figure of 10^5 for basic gain represents the gain of three cascaded stages. Since the overall gain in a computing amplifier must be negative, this represents the next higher figure conveniently attainable; but in solutions of relatively low logon content, such as ours, the reduction in error so obtained would be negligible compared with other errors accepted in the system.

I agree entirely about the risk of over-filling a 4-dimensional display. In almost all computing applications I imagine that some form of selective "strobing" will be required to single out those portions of the picture which are of operational interest.

It is true, as Mr. Cooke-Yarborough suggests, that operation at high speeds does not eliminate all the effects of d.c. drift. It does, however, render negligible the fluctuations occurring during a single solution, which cannot normally be corrected or even spotted. Since we normally measure the operating conditions while, or immediately after, the required solution is on the screen, the chances of error due to drift are much smaller than if we relied on the stability of initial setting-up. We do in fact use chopper-type stabilizing circuits* when conditions demand it, supplying the corrective bias to the cathode of V_1 in Fig. 1. In other units R_K is adjustable; but the normal measuring procedure (Section 7.4) determines input conditions irrespective of d.c. level, and even in the coefficient switching unit shown in Fig. 4(*a*) the initial conditions for any particular solution of the series of 32 can be determined without error from this cause.

The germanium diodes used in the circuit shown in Fig. 4(*b*) (types CG4C and CG6C) were not specially selected; R_1 and R_2 are 47 kilohm resistors, so that our forward current is only 1 mA, and this may account for the small differences in forward voltage observed. It should also be remembered that the input from each switch in Fig. 4(*a*) is attenuated, so that the output steps observed are smaller than the discrepancies produced by S_{1-5}.

If the input, V_c, supplied to the multiplier (Fig. 6) is well stabilized, I do not think that our system need introduce much greater error than one using symmetrical parabolae. In both, we must tolerate the subtraction of large quantities to give small differences; the systems differ mainly in the regions of the (x, y) plane over which the error from this cause is significant.

The concept of accuracy in a computer is potentially as ambiguous as that of noisiness in an amplifier. The user must recognize that the extent to which the apparatus corrupts the information supplied to it will depend on the efficiency with which he utilizes its information capacity. The "percentage of full-scale" in which accuracy normally is expressed is only an indication of this limiting capacity, as Mr. Cooke-Yarborough points out.

In reply to Lt.-Cmdr. Spearman, I would recommend the addition of a drift corrector if the amplifier shown in Fig. 1 were to be used for real-time work or for any problem requiring runs of the order of seconds in duration. It has no special immunity from drift.

My paper was concerned with research into the limitations of

* DEFLEY, E. M.: "The Design of an Electrodynamic Multiplier," *Proceedings I.E.E.*, Monograph No. 87 M, December, 1953 (101, Part IV, p. 187).

techniques rather than with persuading others to use them. It might be most helpful to say first that such computers are not suitable for problems involving many independent variables or those requiring accuracies better than 0·1 per cent. They are at their best with ordinary non-linear differential equations of moderately high order, especially when a large number of relatively short exploratory solutions are required, and parameters or functional relationships have to be flexibly adjusted by trial and error. A good example is offered by eigenvalue problems in atomic physics, where one wants to see how the shape of a potential function affects the eigenvalues of energy. The rapid computation of system response under a wide variety of conditions offers another obvious field. It seems likely too that a computer of this sort could serve as a useful adjunct to a digital machine in many problems, saving time by indicating approximately the region within which the more accurate solution should be sought. But I have not attempted to assess the possibilities in commercial terms.

Mr. Burt, I think, is under a misapprehension. The range of input to the squaring circuits is no greater because v_c is added to $(x \pm y)$. The bias levels, moreover, are chosen so that the output of the squaring amplifier covers its full working range as $(x \pm y)$ varies from maximum to minimum.

The two amplifiers A_2 and A_3, of course, form part of the associated squaring circuits. Had we been prepared to tolerate the disadvantages of simple diode networks, this system would, like Mr. Burt's, have required only three amplifiers. The use of extra feedback amplifiers in the squaring circuits to improve the high-frequency response has nothing to do with our avoidance of symmetrical characteristics; (the same improvement added to Mr. Burt's system would bring his total number of amplifiers up to seven). It is most encouraging to learn that he has achieved such high accuracy.

I share with Mr. Thwaites the vision of a self-programming analogue computer (discussed briefly in Reference 2). As soon as one tries to sketch the outlines of such a machine, however, the interrelation of the necessary digital and analogue processes suggests as the logical conclusion a digital computer with built-in facilities for making analogue computations when required. Between analogue and digital computing processes there is a clear and fundamental distinction, but there seems to be a wide field for the fruitful combination of these processes in the design of analytical engines.

I have found no instability when using the positive-feedback amplifier in a differentiating circuit; but we do not normally use this arrangement because of its susceptibility to overloading by transients in the input. With the silvered-mica condensers used we have had no noticeable trouble from hysteresis.

Mr. Wass correctly emphasizes that this computer is designed for problems other than the continuous simulation of aircraft in flight. We are not, however, restricted to step-function inputs, as Section 5 and References 6 and 7 indicate.

The only novelty claimed for the square-law circuit shown in Fig. 5 is in its use of diodes to control the feedback transfer characteristic of an amplifier, as distinct from other ways of using diodes in function synthesis which are well known.

Prof. Holt-Smith will see from Section 2 that, altogether, 32×32 different trials can be carried out before the system repeats itself. A rate of 12 500 solutions per second is then just sufficient to give a persistent display at about $12\frac{1}{2}$ c/s on a cathode-ray tube, enabling the user to rotate the picture as a whole or to adjust manually a third trial-variable and see its effect on the trends displayed. Where the number of combinations to be tried is smaller, a somewhat lower rate is, of course, advantageous; and as indicated in Section 2, a basic clock rate as low as 800 c/s can be selected for use in such circumstances. Even this speed would be unnecessarily high unless the purpose were to observe the effects of automatic or manual adjustment of parameters or functional relations or angles of projection. For such purposes an intermediate film process would be relatively unsatisfactory. A practical consideration is that the sizes of integrating capacitors can be conveniently smaller at higher rates, and the effects of lower-frequency hum and drift are more readily detected and allowed for.

Instr.-Cmdr. McCleery has made a fair point. To speak of diode switches as "open" when non-conducting is a well-entrenched custom, but I think he is right to dig at it and I wish him success.

534.844 : 621.396.712.3 : 621.395.625.3

Paper No. 1796 R
Feb. 1956

ARTIFICIAL REVERBERATION

By P. E. AXON, O.B.E., M.Sc., Ph.D., C. L. S. GILFORD, M.Sc., F.Inst.P., and D. E. L. SHORTER, B.Sc.(Eng.), Associate Members.

(*The paper was first received* 14*th October*, 1954, *and in revised form* 10*th January*, 1955. *It was published in February*, 1955, *and was read before the* RADIO SECTION 9*th March, and the* NORTH-WESTERN RADIO GROUP 16*th March*, 1955.)

SUMMARY

The demand for a source of artificial reverberation in sound transmission systems is commonly met by the use of reverberation rooms, but attempts have been made to fulfil the requirements by devices involving the transmission of the signal through a number of delay channels. The design and operation of such devices can be treated by simple theory, but in the present state of the art, the quality of artificial reverberation, like that of the natural product, cannot be completely assessed except by ear.

The characteristics of artificial-reverberation apparatus employing delay channels with overall feedback are discussed and illustrated by the results of early experiments using acoustic-delay tubes. The acoustic-tube system, while simple in principle, involves in practice much instrumental complication and becomes uneconomic if good sound quality and long reverberation time are required.

Artificial-reverberation apparatus employing a magnetic-recording delay system, which is free from the defects associated with the acoustic tube, is described. By means of an artifice involving the use of a double recording track, the effective number of delay paths is greatly increased and a high standard of performance is thus attained. This equipment has been in regular use by the B.B.C. for 2½ years.

Artificial-reverberation systems involving a finite number of delay paths tend to produce flutter effects when transmitting impulsive signals. Such effects can be mitigated by the use of a small auxiliary reverberation chamber. By temporarily transferring the signal to the ultrasonic-frequency region, the required effect can be conveniently obtained with a small water-filled reverberation tank. Details are given of equipment of this type which has been designed for use in conjunction with the magnetic recording system referred to.

Part 1. THEORY AND EARLY DEVELOPMENT

(1) INTRODUCTION

Since the earliest days of broadcasting, there has been a demand for some means of supplementing the natural reverberation of the studios. Initially, the chief requirement was for exaggerated reverberation to be used for dramatic effects, and this was obtained from bare-walled reverberation rooms, commonly known as "echo rooms," in which the sound from the studio was reproduced by a loudspeaker. Later, the technique was extended in a more refined form by using an acoustically treated auxiliary studio instead of a reverberation room; this device was employed to introduce a more natural effect in a few special cases where the amount of reverberation at the source of programme was abnormally low. In recent years, a similar requirement has arisen in connection with television, since the acoustic environment in television studios is frequently very "dead," so that for certain types of programme some method of increasing reverberation artificially is essential.

As reverberation rooms are expensive and possess certain technical disadvantages, other methods of producing artificial reverberation have from time to time been proposed. In 1949 various alternatives were considered by the B.B.C. Research Department, and a system using acoustic delay tubes with electrical feedback, for which acceptable performance had been reported in the literature, was selected as the most promising for further study. A series of experiments was carried out which led to a better understanding of the potentialities and limitations of this system and artificial-reverberation systems in general; it became apparent that while relatively crude reverberation effects could be obtained from acoustic-delay tubes without great difficulty, the standard of performance then envisaged could be obtained, if at all, only by bulky and elaborate equipment. Concurrent developments in magnetic recording, however, offered an alternative form of delay element free from many of the difficulties associated with the acoustic system. It was therefore decided to continue the investigation with a magnetic-recording device as the delay element, and the project was carried to the point of designing an artificial-reverberation unit capable of satisfying many of the operational requirements of the service. The first of these units has been in regular operation in television transmissions since July, 1952, while others are now in limited use in sound broadcasting and in the Transcription Service.

In the course of the early experiments it was found advantageous to supplement the delay system by a reverberation chamber having a relatively short reverberation time. As a result of later research this refinement has been provided by the more convenient method of temporarily transferring the greater part of the signal spectrum to ultrasonic frequencies and transmitting it as a pressure wave through water contained in a miniature reverberation tank. Equipment incorporating this device has been in experimental service since April, 1954, as an addition to one of the magnetic-recording units already referred to.

In Part 1 of the paper, the principles common to all forms of artificial reverberation are discussed and the various known devices reviewed. The general theory of delay channels with feedback is given, together with data obtained by experiments with acoustic-delay systems, and the practical limitations of such systems are discussed. In Part 2 the theory is extended to deal with special effects obtainable with magnetic-recording channels and a complete artificial-reverberation unit employing this form of delay is described. Part 3 deals with reverberation phenomena occurring in two- or three-dimensional models, and describes the development of the ultrasonic reverberation tank.

(2) CHARACTERISTICS OF NATURAL REVERBERATION

Before considering devices for imitating the effects of natural reverberation, it may be well to examine briefly the phenomena to be counterfeited and the criteria by which the result is to be judged.

Any source of sound of varying intensity introduced into a room having reflecting walls gives rise to a number of characteristic acoustic effects which are collectively known as reverberation. These effects may be grouped under one of two headings according to whether the response of the room is to be considered as a function of time or of frequency.

The authors are with the British Broadcasting Corporation.

(2.1) Response as Function of Time

If a source of sound is suddenly started, the resulting sound pressure at any given point reaches its steady state in a series of steps, corresponding to the arrival, first of the direct and then of the reflected sound, the latter by a great number of paths of different length. Similarly, when the source of sound is suddenly stopped, the pressure at any point returns to zero in a discontinuous fashion. Because the number of paths by which sound can arrive at a given point is very large, the steps in the growth and decay of reverberant sound are numerous but individually small. The rise and fall in sound level therefore appear to the ear as a continuous process which, in the case of halls and broadcast studios having good acoustics, approximates to an exponential law. In such rooms, even the discrete reflections of impulsive sounds, such as a handclap or pulse of tone, are not separately resolved by the ear.

(2.2) Response as Function of Frequency

At certain frequencies a sound wave reflected from one boundary of a room will arrive back at a given point, after reflection from the other boundaries, in such phase as to reinforce the original, and a state of resonance is thus produced. Every room has a series of such resonance frequencies, at which the level of the reverberant sound is relatively great. The higher-order modes of resonance occur at quite small intervals. The frequencies of the lower modes, however, may be sufficiently far apart to be separately identified by the ear; the results may be subjectively interpreted as an unpleasant timbre or "coloration" imparted to all sounds produced in the room. This effect is most common in small rooms and is an important factor in determining the "goodness" or otherwise of the acoustics.

(3) PRINCIPLES OF ARTIFICIAL REVERBERATION

Artificial reverberation is obtained by modifying the signal in a sound transmission system in a manner imitative of the natural processes described above. The resulting product is then combined with the unmodified signal (with due precautions against unwanted feedback) so as to give an illusion of reverberation occurring at the point of origin. The possible operations that may be carried out on the signal can be divided in principle into one of the following categories:

(3.1) Artificial-Reverberation Systems involving Transmission in Two- or Three-Dimensional Space

The signal is reproduced in a reverberation room or auxiliary studio containing a microphone so placed as to receive a high proportion of reverberant sound; the electrical output of the microphone constitutes the artificial reverberation. Variants on this principle include the substitution of media other than air in the reverberant space and the virtual reduction of this space from three dimensions to two by propagating the sound in a flat box containing a liquid or gaseous medium or in a thin layer of solid material; the use of such devices to supplement other forms of artificial reverberation is discussed in Part 3.

(3.2) Artificial-Reverberation Systems involving Transmission in a Single Dimension

The effect of multiple reflections in a reverberant room is imitated by dividing the incoming signal between a number of parallel channels designed to give different time delays. The attenuation in the individual channels is so adjusted that the delayed signals appearing at the output of the system progressively decrease in amplitude. Each signal may be regarded as an artificial reflection,* and if the series is continued to the point at which further reflections would be inaudible, artificial reverberation can be obtained. To avoid noticeable discontinuities in the growth and decay of sound, the reflections must be closely spaced in time. If each delay channel produced only one reflection, the number of channels would in many cases be inconveniently large. The required end may, however, be achieved more economically by using relatively short delay times and making each signal traverse the delay system repeatedly; an infinite series of reflections can thus be produced. Where transmission can take place in both directions and the attenuation in the medium is sufficiently low, this result may be produced by back-and-forth reflection between the terminations; in other cases, feedback is used, the output signal being amplified and returned to the sending end at a level a little short of that required to produce sustained oscillation.

(3.3) Criteria of Reverberation Quality

The salient feature of natural reverberation is the gradual decay in the sound level after the source has been stopped; the rate of change of energy density in these circumstances, as expressed in the form of reverberation time, is thus the most important single criterion of room acoustics. A mere statement of reverberation time does not, however, give a complete picture of the acoustic properties of a room. The subjective quality of reverberant sound is influenced, in a manner which is as yet imperfectly understood, by the spacing of the resonance modes on the frequency scale or by the spacing of the times of arrival of reflected sound travelling by various paths, both quantities representing different aspects of the geometry of the room. The same considerations apply with even greater force to artificial-reverberation systems, which, while simulating broadly the effects of natural reverberation, do not reproduce the fine structure of the process. It must be emphasized that, in the present state of the art, the quality of reverberant sound, whether natural or artificial, cannot be completely assessed except by ear.

(4) EXISTING ARTIFICIAL-REVERBERATION SYSTEMS

The principles outlined in the last Section are exemplified by a number of artificial-reverberation systems which have been described in the literature.

(4.1) Reverberation Room

The use of reverberation rooms is too well known to require detailed reference. It should be noted, however, that these rooms, although frequently intended to simulate the acoustic properties of large studios or halls, are for economic reasons made relatively small, the required reverberation time being achieved by the use of highly reflecting material for the walls. As already indicated, the quality of the reverberant sound is a function of the geometry of the enclosure, and in fact the reverberation in a small space always has an unpleasant timbre. For this reason, reverberation rooms of practicable size are unsuitable for use in high-quality music transmissions.

The reverberation time of a given room is fixed, and only the ratio of direct to reverberant sound can be readily controlled. Attempts have been made to overcome this defect by dividing the reverberation space into two or even three sections connected be remotely-operated doors,[1] but such expedients add greatly to the cost of an already expensive installation.

(4.2) Mechanical Delay Line

An artificial-reverberation device using mechanical delay lines was developed by Hammond for use with electronic organs,[2,3] and has also been used in film recording.

* For want of a better word, the use of the term "reflection" has been extended for the purpose of the paper to include the individual delayed signals which occur in artificial-reverberation systems.

In this apparatus, the incoming signal is applied to a moving-coil driver unit which sets in vibration a number of wire helices. The reverberant output is obtained from a piezo-electric contact microphone attached to the vibrating system. The vibratory motion is propagated along the helices with relatively low attenuation and reflected back and forth between the various points of discontinuity, giving rise to a series of additional signals at the microphone terminals. The reverberation time of the system is controlled by oil damping of the helices. The device, while possessing the merit of extreme compactness, is subject to certain limitations, to be discussed later, which are inherent in any system having only a small number of delay paths.

(4.3) Acoustic Delay Line with Feedback

The use of acoustic delay lines to give artificial reverberation has been dealt with by Olson[4] in a patent covering, in principle, the use of tubes of different length, each with a loudspeaker at one end and a microphone at the other, or a single tube having a series of microphones mounted on subsidiary tubes branching off at intervals. The delayed signals generated by the various microphones are combined at the output of the system to simulate the effect of a small group of reflections. By the application of overall feedback, the same group of reflections is repeated again and again with progressively decreasing amplitude to form an infinite series.

In a practical embodiment of the above system, described by Curran,[5] the delay paths were arranged to produce a reflection every 23 millisec. Such a series of equally spaced reflections is, of course, an over-simplified substitute for natural reverberation, since it gives only the effect of a single pair of parallel walls. The system described was, however, claimed to give acceptable performance for certain applications in broadcasting.

(4.4) Delay by Magnetic Recording System

Goldsmith[6] and Wolf[2] have described an artificial-reverberation system using a magnetic-tape recorder. The incoming signal was recorded on a continuous loop of tape having 13 reproducing heads followed by an erasing head. The delayed signals generated by the passage of the recorded matter past the successive reproducing heads were combined at the output, thus imitating the effect of a finite series of reflections. The maximum delay provided was half a second, at the end of which time the series came to an end; this effect set a limit to the amount of artificial reverberation that could be used.

(4.5) Delay by Magnetic Recording System with Feedback

A reverberation device specified by Berth-Jones[7] in 1947 contained a magnetic recording system employing a single reproducing head with the addition of feedback, thus giving an infinite series of equally-spaced reflections. In a commercial device advertised in 1950 two reproducing heads were provided, but the combined output from both was fed to the input. Each reflection, on being passed again round the loop produced two more of its kind, the total number being much larger than could be obtained by feedback from a single reproducing head. A similar system was later described[8] utilizing two similar magnetic recording machines running at different tape speeds.

The application of feedback to parallel channels having different delay times, while producing a relatively large number of reflections, leads to some undesirable effects, which are discussed in Sections 6.3 and 13.

(4.6) Delay by Optical Recording on Moving Phosphor

In an artificial-reverberation system described by Goldmark and Hendricks[9] the incoming signal was made to modulate a light source and was optically recorded on the rim of a rotating phosphor-coated disc, a finite series of reflections being produced by a number of photocells arranged around the periphery of the disc. The same series of reflections was repeated with each revolution of the disc, the amplitude diminishing as the glow of the phosphor faded away, and by this means an infinite number of reflections was produced without recourse to feedback. The design appears to have been attended by considerable instrumental complications.

(5) EXPERIMENTS ON ACOUSTIC-DELAY SYSTEMS WITH FEEDBACK

Of the devices reviewed in the last Section, those involving the use of a time-delay system with electrical feedback seem the most promising, if only because the reverberation time can be readily controlled by adjusting the attenuation in the feedback loop.

At the time when the B.B.C. Research Department began to study such systems, the acoustic-transmission tube appeared to be the most convenient means of obtaining the necessary delay. Such a transmission system, while involving elaborate electrical equalization and other instrumental complications, should be capable of operating with little attention for long periods, and has no mechanical parts subject to wear.

A series of experiments with acoustic delay tubes was therefore undertaken; a simple system giving equally spaced reflections was taken as a starting point, and progressively elaborated in the light of subjective tests.

The instrumentation involved in this work is not of great interest for the present purpose, and will not be further considered here. Of greater importance are the general theory underlying the experiments and the principles established therefrom, since these apply also to the design of the equipment described in Part 2.

(6) THEORY OF DELAY CHANNELS WITH FEEDBACK

(6.1) Delay Channel with Single Outlet

It will simplify the discussion if the action of a single delay channel with feedback is first considered in some detail. Such a system is shown diagrammatically in Fig. 1(*a*); the delay channel is illustrated as a loudspeaker, acoustic-delay tube and microphone, together with such amplifiers and equalizers as are necessary to give zero gain over the working frequency band, but the ensuing discussion applies equally to a recording channel or to any other transmission medium having a constant time lag between input and output. For simplicity, it is assumed that any phase shift in the amplifiers, equalizers, loudspeaker and microphone is negligible compared with that occurring in the tube itself, and furthermore, that the feedback is positive in sign.

Consider first the relationship between the reverberation time, delay time and attenuation in the feedback loop, the last-named quantity representing the margin of safety of the system against self-oscillation.

Let t be the time delay in seconds, and T the reverberation time, i.e. the time after which a signal passing repeatedly round the feedback loop falls by 60 dB. In the course of time T, the signal must traverse the loop T/t times, suffering each time an attenuation of A decibels, whence

$$A = 60t/T \text{ decibels, or } T = 60t/A \text{ seconds.}$$

To simulate natural conditions it may be required to make the reverberation time vary with frequency, and to this end, the response of the feedback circuit may be made frequency-dependent by the insertion of an appropriate network. It should be noted that when A is small, as must be the case if the delay time required is to be kept to a minimum, slight changes in the response of the feedback loop cause large changes in

reverberation time. For example, with a 50 millisec delay, for which the path length in air would have to be about 55 ft, and a reverberation time of 1·5 sec, the feedback loop must be adjusted to have an attenuation of only 2 dB. A variation of ± 1 dB in the loop response will then cause the reverberation time to vary between 3 and 1 sec, while the system will oscillate if the response rises by 2 dB at any of the frequencies (occurring at intervals of $1/t$, or 20 c/s throughout the band) at which the loop phase shift is zero. It will be seen that there is little latitude for unintentional variations in the loop attenuation and that stringent requirements are imposed on the frequency response of the delay element.

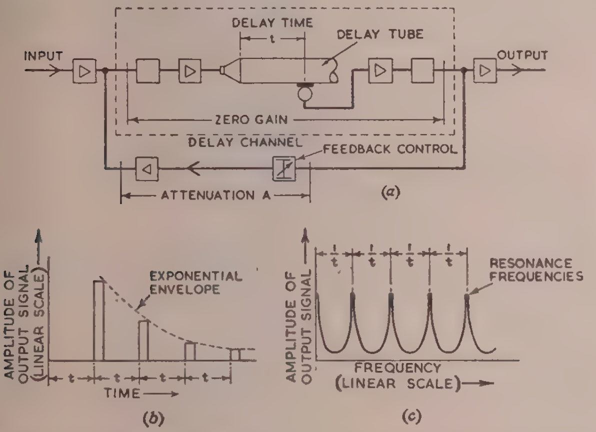


Fig. 1.—Artificial-reverberation system using single delay channel with feedback.

(*a*) Simplified schematic.
(*b*) Response as a function of time. Pulse input.
(*c*) Response as a function of frequency. Tone input.

Key to graphical symbols used in all the Figures.

Amplifier. Equalizer.

Attenuator. Loudspeaker

Microphone.

Consider next the response of the system as a function of time. Fig. 1(*b*) shows the output resulting from a transient signal, such as a burst of tone, entering the circuit at zero time. This output consists of a series of reflections of the original signal (the envelope amplitude only being shown), occurring at equal time intervals t, each reflection being attenuated by A decibels with respect to the previous one, so that the overall envelope represents an exponential decay. The response of the system superficially resembles that of a reverberant room but differs from it in two respects:

(*a*) With the large values of t likely to occur in practical artificial-reverberation systems, the silent interval between individual reflections may be perceptible to the ear, giving an impression of flutter.

(*b*) The series of reflections can be regarded, apart from its exponential decay, as a repetitive phenomenon having fundamental frequency $1/t$ plus a series of harmonics. Thus, even when t is too short for flutter effects to be recognized, the system may still be capable of giving the subjective impression of an assignable pitch.

It will be seen, in fact, that the circuit of Fig. 1(*a*) reproduces all the phenomena associated with reflections between two parallel walls.

Consider finally the overall frequency characteristic of the system. The response passes through a series of maxima, as shown in Fig. 1(*c*). Each peak represents a condition in which the signal fed back from the output of the delay channel is in phase with, and thus reinforces, the incoming signal, a state of affairs which can readily be shown to occur at frequency intervals of $1/t$. Here again, the response of the system has a superficial resemblance to that of a reverberant room, with its series of resonance modes; in the neighbourhood of each "mode" in Fig. 1(*c*) the circuit, in fact, exhibits the characteristic resonance phenomena, such as beats between free and forced oscillations on the sudden application of a signal having a frequency slightly different from the frequency of maximum response. In an artificial-reverberation system, however, the value of t is likely to be such as to place the modes much further apart in frequency than those of a reverberant room. The reinforcement of individual components of a complex waveform, such as that of speech or music, may then be noticeable to the ear, the effect being known as "coloration"; the situation is aggravated by the fact that the peaks in Fig. 1(*c*) form a harmonic series.

The time and frequency response described in the last two paragraphs are not, of course, independent, but are alternative aspects of the same* performance. The close relationship between the two can be illustrated if the incoming pulse of Fig. 1(*b*) is considered to be applied to a circuit having the frequency response of Fig. 1(*c*); the resulting impression of pitch could then be attributed to the shock excitation of the series of harmonically related resonance modes.

For more complicated systems involving several parallel paths giving different delay times, the relationship between the time and frequency response is of a more complex character. Moreover, the ear recognizes in all cases a fairly clear distinction between time phenomena of the flutter class and frequency phenomena involving the imposition of a characteristic timbre or coloration—of which Figs. 1(*b*) and 1(*c*) illustrate the simplest form. It is therefore convenient to examine the performance of artificial-reverberation systems directly from either or both points of view.

(6.2) Delay Channel with more than One Outlet: Feedback over Single Path

To increase the number of reflections, the incoming signal may be made to follow a number of parallel paths of different lengths. The same effect may, however, be achieved with greater economy by a single delay channel having outlets at a number of points along its length. Fig. 2 shows how the simple system of Fig. 1

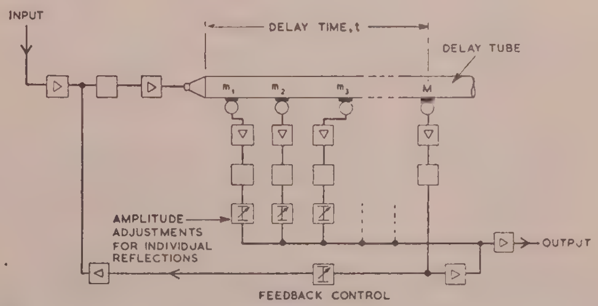


Fig. 2.—Artificial-reverberation system using single delay channel with more than one outlet. Feedback over single path.

can be elaborated in this way by employing four microphones m_1, m_2, m_3 and M, spaced along an acoustic delay tube. It should be noted that the feedback is taken from the last microphone M only; signals from the intermediate microphones m_1, m_2

* To give a complete description of the system, Fig. 1(*c*) would need to be supplemented by a phase/frequency curve.

and m_3, while contributing to the output of the system, do not circulate in the feedback loop.

If the four microphones were equally spaced along the tube, the system of Fig. 2 would produce a series of reflections at equal time intervals $t/4$. As has already been pointed out, however, a series of equally spaced reflections may produce unwanted effects, and it is therefore desirable that the microphones should be placed at irregular intervals along the tube. The effect of this artifice will be seen from Fig. 3(*a*), which shows the amplitude and spacing of the reflections produced by the system of Fig. 2; here the amplitude is plotted on a logarithmic scale, so that an exponential-decay envelope would appear as a straight line, and the individual reflections appearing in the output of the system are marked according to the microphone by which they are picked up. It will be seen that the irregular spacing of the microphones does not entirely remove the periodic element from the series, for the time pattern of the group of reflections generated by these microphones is repeated at each passage of the signal round the feedback loop. The subjective effect of this recurrent pattern will be referred to as "group repetition flutter."

Successive signals from microphone M, being attenuated by a constant fraction at each passage round the feedback loop, decrease exponentially in amplitude. The signals from m_1, m_2, and m_3, however, have to be adjusted in amplitude, by regulation of the individual attenuators shown, to produce the smooth decay envelope of Fig. 3(*a*); if this is not done and conditions as shown in Fig. 3(*b*) obtain, the artificial reverberation becomes amplitude-modulated at the group repetition frequency $1/t$, and the corresponding flutter effect is considerably aggravated. For the same

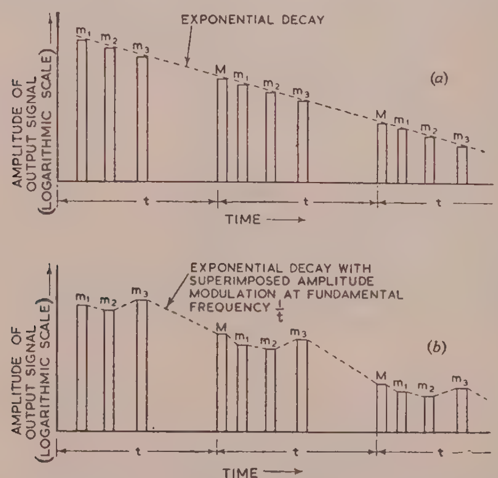


Fig. 3.—Amplitude and spacing of reflections produced by a single pulse applied to the artificial-reverberation system of Fig. 2 with (*a*) correct, and (*b*) incorrect adjustment of the individual reflection controls.

reason, any change in the reverberation time requires a readjustment of the gain controls associated with m_1, m_2 and m_3, and if the system is designed to have a reverberation time varying with frequency the attenuation in the individual microphone circuits will have to be made to vary appropriately with frequency if the exponential-decay envelope is to be preserved.

The steady-state response of the multi-path system of Fig. 2, like that of the simple system of Fig. 1, passes through a series of maxima at frequency intervals $1/t$. Additional maxima, however, appear at frequencies at which the outputs of two or more microphones are in phase and thus reinforce the signal. If the microphones are irregularly spaced, these frequencies of reinforcement are irregularly spaced also; moreover, the peaks in the frequency response curve differ in height, according to the number of microphone outputs in phase. The multi-path system, with its irregularly spaced "modes," thus gives a better approximation to the response of a reverberant room. There remains, however, the possibility that one of these modes may be so far separated in frequency or amplitude from its neighbours as to lead to a conspicuous coloration. Such effects cannot be entirely avoided, but can be minimized by appropriate spacing of the microphones. The determination of the best condition involves a detailed study of the relationship between the spacing and the wavelength of the signal in the delay medium and will not be considered here; the problem is fully discussed in Part 2 in relation to the spacing of pick-up heads in the magnetic recording system.

(6.3) Delay Channel with more than One Outlet: Feedback Over Parallel Paths

In the circuit of Fig. 2, the number of reflections produced would clearly have been increased if, in addition to the output from M, the outputs from m_1, m_2 and m_3 had been fed back to the input. However, such an arrangement is in general impracticable, since it involves the application of feedback to a circuit which consists of several parallel paths of different lengths and which must therefore have already an irregular frequency characteristic. The points of maximum response of the feedback loop will occur at those few frequencies at which all the microphone outputs are in phase, and when feedback is applied, reverberation will be practically confined to the vicinity of these points.

A border-line case arises with a circuit such as that of Fig. 4, in which the number of parallel paths in the feedback loop is only two. Here, the loop response reaches a fixed maximum

Fig. 4.—Artificial-reverberation system with feedback over two parallel paths.

value—the sum of the responses of the two microphones M_1 and M_2—at a series of frequencies extending throughout the working band and having a constant separation $1/\tau$, where τ is the difference in time delay between the two. At each of these frequencies the reverberation time obtained by applying feedback will be the same, and the system, while subject to the disadvantage of having regularly spaced modes, is therefore workable.

It may be of interest to examine the response of the circuit in Fig. 4 as a function of time; it is assumed that both M_1 and M_2 are connected to give positive feedback. A single short tone pulse, arriving in succession at M_1 and M_2, produces two pulses separated by a time interval τ. These, on retraversing the delay medium, produce successively a pulse from M_1 alone, a pulse from M_1 and M_2 acting simultaneously, and a pulse from M_2 alone; the second of these pulses, being the sum of two others,

has the largest amplitude. Continuation of the process produces a complicated effect which is best illustrated by an example. Fig. 5(*a*) shows the decay process following the arrival, at zero time, of a single tone pulse; as in Figs. 1 and 3, only the envelope amplitude is shown. For the purpose of this example, it is assumed that the attenuations in the feedback loop, when the output is taken from M_1 or M_2 alone, are 8 dB and 11 dB respectively, whence it can readily be shown that the margin of safety against self-oscillation is 3·3 dB. It will be seen that the decay envelope shows a high degree of amplitude modulation; this may be expected to appear as a flutter effect and is a further disadvantage of the arrangement. For comparison, Fig. 5(*b*) shows the series of reflections which would be obtained using the same delay tube and microphones and with the same safety margin, but with feedback taken from M_2 alone, the output of M_1 being connected in a similar way to that of m_1, m_2, and m_3 in Fig. 2 and adjusted to fit the exponential-decay envelope. It should be noted that the wide gaps between the reflections in Fig. 5(*b*) could be filled, as in Fig. 2, by the addition of further microphones not included in the feedback loop, but that the same expedient applied to the circuit of Fig. 4 could not produce a smooth decay.

No precise value of reverberation time can be assigned to Fig. 5(*a*), but even if the largest reflections only are counted it is clear that the rate of decay is much greater than in Fig. 5(*b*).

From the foregoing it may be concluded that feedback over even two different paths produces undesirable effects, and a single feedback path is therefore to be preferred.

(7) EXPERIMENTAL RESULTS

It is not proposed to describe in detail the delay-tube experiments already referred to. The main steps in the investigation and the conclusions reached are summarized in the following Sections.

(7.1) Simple System giving Equally Spaced Reflections

The subjective effect produced by the simplest types of artificial-reverberation system, giving regularly recurring reflections, was investigated. When the interval between reflections exceeded some 25 millisec, the subjective result could be described as a flutter echo. As the interval was reduced, there appeared in addition an effect referred to by observers as a "honk" or "twang," giving an impression of definite pitch and imparting a characteristic coloration to the signal. At 10 millisec spacing, the flutter effect was negligible and the coloration effect was predominant. Of the two phenomena, the flutter was considered slightly less objectionable, but the amount of artificial reverberation that could be added to the original signal without one or other of these defects becoming evident was very small.

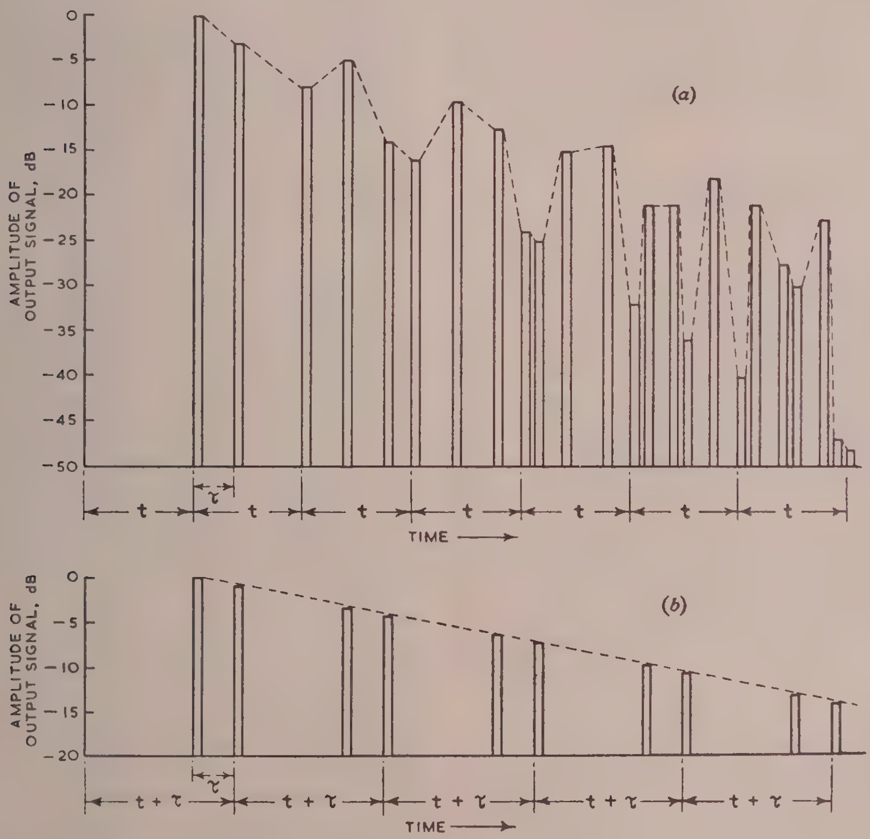


Fig. 5.—Amplitude and spacing of reflections produced by a single pulse applied to

(*a*) Circuit of Fig. 4. (*b*) Circuit modified by taking feedback from M_2 alone.

(7.2) Combination of Three Simple Systems

To simulate more closely the conditions obtaining in a reverberant room, three independent artificial-reverberation circuits, each similar to that of Fig. 1, were combined, thus imitating the natural reflections which occur between opposing pairs of side walls and between floor and ceiling. Such a combined system has three series of resonance modes, which correspond to the axial* modes of a room having dimensions equal to half the lengths of the three delay tubes used. The remaining modes of the room, involving transmission of sound along oblique paths between adjacent walls, are not, of course, reproduced by this method, but it was hoped that a satisfactory approximation to natural reverberation might be achieved without them. In fact, this hope was not realized; the flutter and coloration effects associated with the individual systems were still evident to the ear. This result is of some interest as a sidelight on the theory of room acoustics, since it suggests that the axial modes alone are an inadequate criterion of performance; the point is further discussed in Part 3.

From the foregoing it was concluded that satisfactory artificial-reverberation effects could not be obtained by a system giving equally spaced reflections.

(7.3) Multi-Path Systems

A series of experiments was next carried out on multi-path systems similar to that of Fig. 2 in order to determine the minimum overall time delay t and the maximum interval between reflections which would give tolerable reverberation quality.

For a given reverberation time, t is proportional to the loop attenuation A. The latter quantity is governed in turn by the gain stability of the feedback loop and the smoothness or otherwise of its frequency characteristic; in a practical acoustic-delay system, its minimum value may be 3–4dB. Apart from these considerations, however, a large value of t is desirable on account of the group repetition flutter, which was found to become progressively less objectionable as the frequency $1/t$ was lowered.

The upper limit to the mean interval between successive reflections appeared to be in the neighbourhood of 30millisec, and at least four microphones had to be employed to allow the necessary degree of irregularity in their spacing.

The following data, relating to two of the experimental systems employed, serve as an example of practical operating conditions:

System A

Overall delay time	50millisec
Frequency of group repetition flutter	20c/s
Number of delay paths	4
Mean interval between reflections	12½millisec
Frequency range	50–7 000c/s
Reverberation time at 100c/s	1sec
Reverberation time at 4 000c/s	½sec

System B

Overall delay time	135millisec
Frequency of group repetition flutter	7c/s
Number of delay paths	5
Mean interval between reflections	27millisec
Frequency range	50–4 000c/s
Reverberation time at 100c/s	2sec
Reverberation time at 4 000c/s	1sec

System A roughly simulates the acoustics of a small broadcasting studio, but has too short a reverberation time to be of much use for musical programmes.

In system B the increased delay time is accompanied, for reasons given later, by a reduction in bandwidth. Flutter effects were still noticeable at high levels of reverberation, particularly with impulsive signals. The quality was nevertheless considered sufficiently good for use with certain types of music.

* The axial modes of a room are those involving transmission of sound in directions normal to the boundary surfaces.

(7.4) Variation of Reverberation Time and Amount of Reverberation with Frequency

In all cases a more natural effect was obtained by making both the reverberation time and the amount of reverberation fall with rising frequency. These measures may be regarded as simulating the natural increase with frequency of the absorption in rooms and of the directivity of sound sources; however, they served also to mitigate the defects of the artificial reverberation.

(7.5) Use of Auxiliary Reverberation Chamber

Flutter effects in the artificial-reverberation systems described were appreciably reduced by transmitting the incoming (or outgoing) signal through a loudspeaker and microphone placed in a small reverberation chamber; the reverberation in this chamber served to fill the intervals between reflections and did not therefore require to be very long. The improvement was, however, obtained at the cost of introducing the unpleasant coloration which is a characteristic of small rooms, and for this reason the use of an auxiliary reverberation chamber did not become a practical proposition until the development of the ultrasonic tank described in Part 3.

(8) FURTHER DEVELOPMENT

The reverberation time so far attained was still insufficient for many purposes, and a further increase in delay time was therefore desired. Unfortunately the delay time which can be achieved by an acoustic tube is severely limited by the heavy attenuation suffered by the transmitted sound. For lengths up to 100ft, giving delay times of about 90millisec in air, tubes of 1in bore may be employed to transmit frequencies up to about 7 000c/s. For greater lengths it may be necessary to reduce the attenuation by increasing the bore and to tolerate some limitation of the upper frequency range through transverse resonance effects. While a certain deficiency at high frequencies is a common feature of natural reverberation, and may therefore be tolerated in the artificial product, an increase in tube diameter involves a serious increase in the space taken up by the equipment. An example of this is the experimental system B, in which a tube of 1½in bore was used to transmit frequencies up to 4 000c/s. The tube had an overall length of 165ft, of which the last 14ft was filled with sound-absorbent material to minimize reflection; when mounted in the form of a coil it occupied some 70ft^3. To obtain a greater delay time without further reduction in bandwidth, two or more complete channels could be connected in cascade, but apart from the cost of the extra loudspeakers, microphones, equalizers and amplifiers, the equipment would then require for its accommodation a room comparable in size with the reverberation chamber which it was intended to replace.

It was clear, therefore, that any further development would require the use of a more compact delay medium, and the possibilities of magnetic recording were reconsidered. This form of delay had hitherto been open to some practical objections on account of the wear on the tape and heads, which set a limit to the length of time that the equipment could be operated without maintenance. Fortunately, later developments in magnetic recording had made it possible at this stage to construct a delay channel with the heads working out of contact with the magnetic medium. It was therefore decided to construct a magnetic-recording machine for producing artificial reverberation. The construction and operation of this machine are described in Part 2.

Part 2. THE MAGNETIC-RECORDING ARTIFICIAL-REVERBERATION MACHINE

(9) INTRODUCTION

A description will be given in this Part of the paper of an artificial-reverberation apparatus which employs a magnetic-recording delay system. Fig. 6 is an external view of the delay unit in which the programme signals are recorded on the inside rim of a basin-shaped wheel, which is coated with magnetic material, and are reproduced at subsequent intervals by a series of reproducing heads spaced round the rim. With suitable spacing of the reproducing heads, and by suitable attenuation of the signals reproduced from them, it has been possible to achieve a satisfactory simulation of the decay of sound in a reverberant enclosure. The general principle by means of which the same sound is made to traverse the delay chain repeatedly, until its reproduced intensity reaches noise level, has been explained in Part 1. For reasons of space, flexibility and economy, a similar feedback system is used in the magnetic-delay unit, and every signal appearing in the last reproducing head is fed back into the recording chain for re-recording on the wheel rim. The basic delay in the magnetic system can, without inconvenience, be made greater than in the acoustic-tube system, and in the apparatus to be described it is some 250 millisec, i.e. the group repetition rate is about four per second.

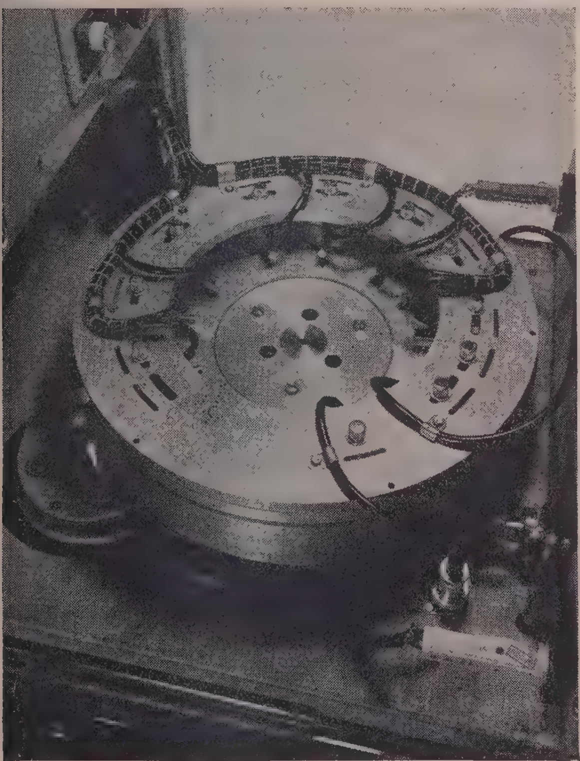


Fig. 6.—Magnetic-recording delay unit of artificial-reverberation equipment.

The rate of decay and the shape of the decay pattern associated with each signal fed into the apparatus depend on the relative attenuation of the outputs from the successive reproducing heads and on the level at which the signal is re-recorded from the feedback chain. It is possible to obtain a series of reverberation times by suitable choice of these factors, the upper limit being determined by the inherent liability of the feedback system to go into oscillation.

(10) THE MULTI-TRACK SYSTEM

It is desirable that the number of artificial reflections generated in a given time should be as high as possible, otherwise when the apparatus is fed with an impulsive sound its output acquires an unpleasant "chopped" character owing to amplitude flutter. A difficulty arises here with the finite dimensions of reproducing heads, which set a limit to the minimum spacing possible between the reproducing gaps of adjacent heads. The time interval provided by this spacing may, of course, be made shorter by increasing the speed at which the recording medium passes the heads. However, this cannot be carried too far without serious inconvenience, for each such increase of speed entails a reduction of the overall delay available. Consequently the level of signal which must be fed back to the input of the system from the feedback head becomes higher, and, as shown in Part 1, this imposes increasingly stringent tolerances on the accuracy to which the system must be equalized if self-oscillation, or coloration, is to be avoided. A compromise must therefore be made between the factors of head spacing, recording speed and feedback level.

A reasonably numerous and complex occurrence of reflections, with acceptable values of feedback level and recording speed, has been obtained by the use of a multi-track recording system.[10] This can simulate the effect of a multitude of reproducing heads whilst requiring the physical presence of only a fraction of that number of heads. Fig. 7 illustrates the basic means by which the effect of many reproducing heads may be obtained. A simultaneous multi-track recording of two or more tracks is made in the magnetic medium, the various recording heads being displaced longitudinally with respect to one another so that at some subsequent reproducing head, which spans all the tracks and reproduces them simultaneously, there is a time interval between the reproduction of corresponding parts of the programme on each track. Suppose that A, B and C are three recording heads which are recording the same programme on magnetic tracks, *a*, *b* and *c*, whilst S, T, U, V, etc., are reproducing heads which can reproduce the recorded signals from the three tracks simultaneously. The longitudinal displacement of the tracks *a*, *b* and *c*, when reproduced by head S, introduces differing time lags in the reproduction of any given signal fed to the apparatus, the recorded signal of *c* being reproduced first and that of *a* last. If the tracks *c*, *b* and *a* are recorded at a progressively lower level, the output of the head S is identical to that which would be obtained using a single track (with one recording head) reproduced by three displaced reproducing heads, the outputs of which are progressively attenuated. When more than one reproducing head is introduced into the system the advantages of this multi-track method become evident. Three recording heads (A, B and C), two reproducing heads (S and T) and three tracks will produce six separated responses for the use of five heads. A single-track recording would require seven heads to attain the same result. Similarly, three recording heads (A, B and C), three reproducing heads (S, T and U) and three tracks will produce nine responses separated in time, a result which the single-track recording would require ten heads to accomplish. The more reproducing heads which are employed in the system, the greater is the relative number of responses obtained by multi-

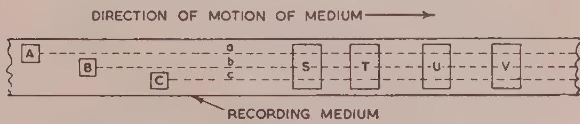


Fig. 7.—Multi-track recording and reproduction.

A, B, C	Recording heads.
a, b, c	Recorded tracks.
S, T, U, V	Reproducing heads.

track recording. Increasing the effective number of reproducing heads in this manner, and using suitable feedback arrangements, enables a reverberation pattern to be produced which meets most normal requirements except those involving impulsive sounds such as pistol shots.

(11) THE DELAY UNIT

The mechanical arrangement of the recording and reproducing unit which provides the delayed reflections is shown in Fig. 8. In the present apparatus the decay pattern is provided by the use

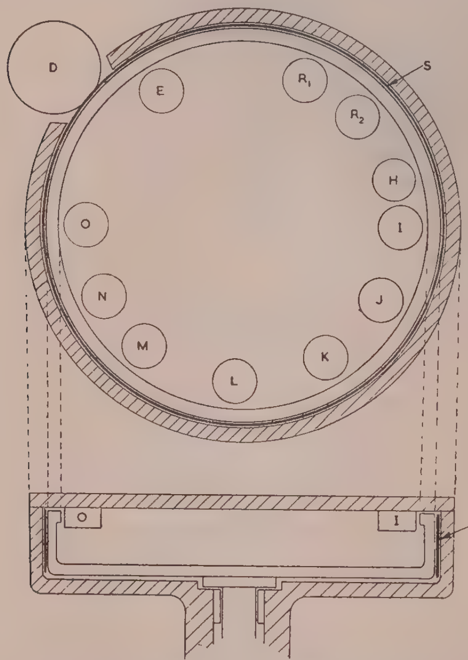


Fig. 8.—Mechanical arrangement of magnetic-recording delay unit.

of two recording heads (and tracks) and eight reproducing heads. The magnetic coating is sprayed, in the form of an endless track ¼in wide, on the inner edge of the basin-shaped rotating member. This basin, which has an inner diameter of 9½in, is supported and enclosed in a bearing unit and is rim-driven by the driving wheel D which is fixed on the shaft of a synchronous motor (not shown). The basin is rotated at about 3r.p.s. and the magnetic medium is thus travelling at a speed of approximately 100in/sec. The various erasing, recording and reproducing heads are hung from a plate into the well of the basin so that their gaps face outwards on to the magnetic medium. With this arrangement the whole system can be well screened magnetically, and it can also be made fairly dustproof. To avoid wear, and the necessity of frequent renewal of the magnetic track, the various heads work out of contact with the medium, and are separated by approximately 0·001in from it. This separation must be maintained fairly accurately in the normal range of temperature variation. Accordingly, the basin-shaped rotating member and its enclosure (which supports the heads) are made of the same non-ferrous alloy, so that the radial movements of the track and heads, when the structure expands or contracts, are for practical purposes identical.

R_1 and R_2 are the two half-track recording heads which record the programme on the upper and lower halves of the magnetic track respectively, whilst H, I . . . N, are reproducing heads of conventional design which reproduce the signals from both tracks simultaneously. The output from the reproducing heads is suitably attenuated, as described previously, and the relative displacement of the recording heads, R_1 and R_2, doubles the number of reproduced reflections in the time interval of 250millisec in which any part of the recorded track passes from R_1 to the last reproducing head.

The head O provides feedback, through suitable amplifiers, to both of the half-track recording heads, but it is made to reproduce, and provide feedback from, one track only. Taking feedback from both tracks through a common head would introduce another source of amplitude flutter into the output of the apparatus. This is an instance of flutter arising when feedback is taken over two parallel paths, which was discussed in general terms in Part 1. The mechanism of the effect in the present apparatus is described in more detail in Section 13.

The head E is an erasing head which removes the signal from the medium, leaving it neutral and ready for a new signal to be recorded by the heads R_1 and R_2. All the heads are suitably screened in Mumetal boxes in order to avoid hum and crosstalk, and the outside of the basin is screened by a Mumetal ring S (except where it is in contact with the driving wheel D) to reduce pick-up from motor-drive and power units.

(12) ELECTRONIC EQUIPMENT

(12.1) General Scheme

A block schematic of the arrangement of the electronic equipment necessary to enable the delay unit to fulfil the functions previously described is given in Fig. 9. The programme, which may come direct from the microphone chain or from an existing

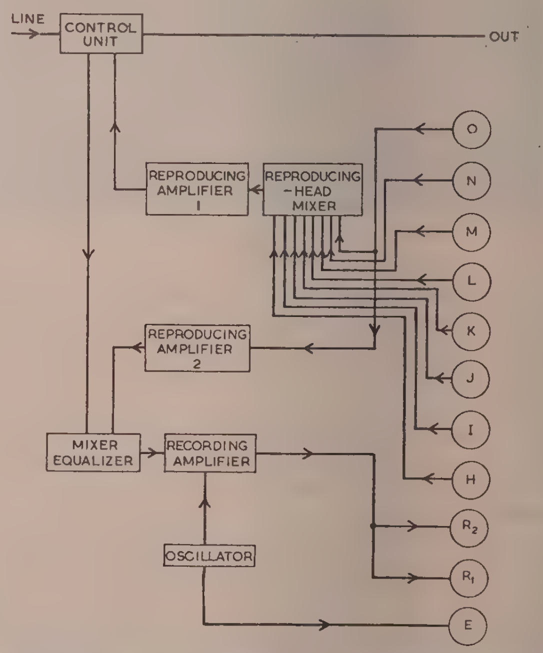


Fig. 9.—Simplified block schematic of artificial-reverberation equipment.

recording, passes into a control unit from whence it proceeds in two parallel paths, one straight through to the transmitter or other terminating equipment, and the second to the artificial-reverberation side chain. From the mixer equalizer, which is the first unit in the side chain, the programme passes to the recording amplifier with which is associated the normal high-frequency

oscillator to supply bias current. The same oscillator also supplies the erase head with the required erasing current. The recording-amplifier output passes into the half-track recording heads, and the programme is then recorded on the two magnetic tracks. The outputs of all eight reproducing heads are fed into a reproducing-head mixer which attenuates the separate outputs in the desired manner to reproduce a suitable decay curve. After suitable amplification in the reproducing amplifier, the output of the reproducing-head mixer passes once more into the control unit to join the original programme line for feeding to the transmitter or other terminating equipment. The final head in the reproducing chain (the feedback head) also feeds, through another reproducing amplifier, into the mixer equalizer so that it is recorded again at suitable level to continue the decay curve in the manner previously described.

(12.2) Amplifiers

The recording amplifier is a conventional two-stage amplifier which uses current feedback to maintain a constant-current/frequency characteristic in the recording heads.

The two reproducing amplifiers employed in the apparatus are identical and follow conventional magnetic-recording practice. Special precautions have been taken against hum and microphony.

(12.3) The Oscillator

The oscillator, which supplies both bias and erase currents, is of the Wien-bridge type and feeds two pairs of valves in push-pull, one pair being associated with the recording heads and the other with the erase heads. High levels are required in the bias and erase signals, and attention has been paid to the reduction of even-harmonic distortion in order to obtain a satisfactory noise level. Final reduction of noise due to imperfect oscillator waveform is carried out using a small rectified current fed to the recording and erase heads in the manner suggested by Stott.[11]

(12.4) Reproducing-Head Mixer Unit

It has been explained, in general terms, that the reverberation time can be controlled by suitable attenuation of the individual outputs from each successive reproducing head and from the final feedback head. In laboratory equipment, suitable voltage dividers may be provided to enable the attenuation introduced in each output to be continuously variable so that any required rate or shape of decay may be obtained. With equipment for operational purposes, however, the careful setting-up which would be required becomes impracticable, and the potentiometers are replaced by a wafer switch containing banks of fixed resistors. The wafer switch installed in the present apparatus has ten separate positions, and attenuating resistors are fitted so as to provide nine preset reverberation times. The function of the tenth position is described in Section 12.6. The attenuating resistors must be carefully adjusted so that in each case the fall in the amplitude of the signal reproduced from successive heads follows a smooth exponential curve of the type shown in Fig. 3(*a*). Faulty adjustment of these resistors will lead to flutter effects as illustrated in Fig. 3(*b*). The reverberation times available range from 0·6 to 5 sec.

(12.5) The Control Unit

When conventional echo-mixing facilities are available, the apparatus may be fitted into a normal studio control system in precisely the same way as a reverberation room. The control unit is employed with the artificial reverberation machine when such facilities are not available. To obtain the maximum signal/noise ratio, the mixing of the direct and reverberant signals is carried out at the output of the machine, and by the use of suitable attenuators (and/or faders) the ratio of direct to reverberant sound may be varied over wide limits, in conjunction with the available reverberation times on the machine itself, to provide the desired subjective effect. Therefore, although the artificial-reverberation machine may be employed directly in place of a reverberation room in studio control equipment, it can, in addition, provide a variable reverberation time which a normal reverberation room does not.

(12.6) Monitoring of Individual Head Outputs

The use of a wafer switch to provide nine predetermined reverberation times has been described earlier. The tenth position on this switch, in conjunction with a second rotary switch adjacent to the delay unit, enables the output of any individual head to be monitored separately. In this tenth position of the switch the attenuating resistors in the output circuits of the various heads are short-circuited and the individual outputs become equal. The second switch within the apparatus then enables any one particular head to be monitored. This facility is useful in the setting-up and checking of the machine and also enables straightforward "echo," as opposed to "reverberation," effects to be obtained from the machine. The eight reproducing heads provide eight set delay times, up to a maximum of about 250 millisec. The studio control facilities, or the control unit with which the apparatus is used, enable the ratio of direct sound to echo intensity to be set up as required.

(13) THE OCCURRENCE OF AMPLITUDE FLUTTER

The amplitude flutter which can occur owing to the finite spacing between reproducing heads has been briefly mentioned in describing the multi-track recording system. A very sharp impulse of sound, occurring with little or no background of other sound, appears in the output of the apparatus with some of the qualities of reverberation but with a "chopped" quality which lacks naturalness. This occurs when there is a period of actual silence between the signals reproduced by successive reproducing heads. In a continuous type of sound, or when an impulse is heard against a fairly high-level background of other continuous sounds, the effect may not be apparent at all because the actual amplitude changes involved are of much lower order. The effect also becomes less unpleasant as the length of the impulse is increased, and when this length becomes larger than the maximum spacing between two consecutive heads, the flutter disappears. This is because the sound fills in the maximum time interval between any two consecutive responses and there is no break in the output from the apparatus. For the same reason the flutter disappears when dealing with sounds whose frequency spectrum lies below a critical frequency which has a wavelength in the recording medium equal to the maximum spacing between two consecutive heads. Not unexpectedly, in view of the last result, subjective experiments also reveal that the flutter is reduced if the impulse has a gently sloping build-up and decay.

The mean spacing of the reproducing heads, in conjunction with the multi-track system, is thus the critical factor in avoiding the more obvious subjective effects of this amplitude flutter. In the present delay unit the mean reproducing-head spacing represents a time delay of about 30 millisec, so that in the two-track system a short recorded impulse will be reproduced at mean intervals of about 15 millisec. In practice it has been found that the earlier reproducing heads, such as H, I and J, which are producing the highest level of output, should be closer together than the mean spacing. If they are too widely spaced the flutter associated with the disappearance of the signal is both of large amplitude and low frequency and becomes very noticeable. This restriction is less important in the later heads which are repro-

ducing at lower level. It is of some interest that the first traversal of the delay chain, although it occupies only a total of 250 millisec, is so critical in this matter.

The effect on impulsive sounds is also less obvious if the sound is originating from an enclosure which possesses some reverberation of its own, when the sharp edges of the pulse tend to be blurred before reaching the artificial-reverberation machine, and the periods of silence between the outputs from successive heads tend to get filled up. This is in conformity with the observations that the "chopped" effect is less noticeable when the impulse is rounded off or is heard against a continuous background of sound. It may also be demonstrated, as indicated in Part 1, by first feeding the impulsive sounds originating in a "dead" enclosure into a small studio of comparatively short reverberation time, which provides partial diffusion, before passing them to the reverberation machine to obtain the much longer reverberation times of which it is capable.

Amplitude flutter will also arise from any arrangement of the feedback system which corresponds to feedback over parallel paths of different length. It may arise, for example, if the head O feeds back the signal from both tracks, and in this case the two parallel paths are from the recording head R_1 to the feedback head and from the recording head R_2 to the feedback head. If a sharp impulse of sound is fed into the apparatus it is recorded on each track, and the subsequent feedback results in four more impulses. It may be shown, as in Section 6.3, that the two middle impulses of this set are directly opposite to one another on the magnetic tracks, so that they will add, when reproduced, to give double the intensity of the impulses on either side. The effect is cumulative, so that when these impulses are fed back for the second and subsequent times an increasing series of evenly-spaced impulses, the amplitude of which rises and falls, is produced and there is introduced a noticeable amount of amplitude flutter. This is eliminated if the feedback head O reproduces from one track only, although its output may be fed back into both recording heads. In this connection it is interesting to observe that at recording speeds of the present order, which result in very long wavelengths, the spread of flux from the upper half-track is still sufficient to induce a noticeable signal into the half-track feedback head O facing the lower track. A Mumetal strip must then be placed immediately above the single-track feedback head to divert the unwanted flux from the upper track into a path away from the feedback-head core.

(14) THE OCCURRENCE OF FREQUENCY COLORATION

(14.1) Reproducing-Head Spacing and Coloration

The reproducing-head spacing is, as shown, partly determined by the requirement that flutter in the apparatus shall not reach objectionable proportions. However, the relative spacing can also have a marked influence on the occurrence of frequency colorations. The distance between any pair of heads must correspond to some recorded wavelength and to a multiple of the harmonically-related shorter wavelengths. When the frequency corresponding to any of these wavelengths is being fed into the apparatus the two heads concerned will be reproducing exactly in phase, thus giving a maximum combined amplitude which depends on the relative attenuations in their output circuits, i.e. on the reverberation time chosen. Minima in the response will exist, also, when the distance between any pair of heads is an odd multiple of a half-wavelength of the frequency being recorded. In fact, when any frequency is fed into the apparatus the resultant output at some instant later can be represented by the vector sum of a set of vectors, each of which represents in phase and magnitude the output of one reproducing head. The magnitude of each vector, with respect to the other vectors in the set, will depend on the reverberation time chosen and the position of the head along the delay chain. The phase of each vector will depend both on the position of the head and the recorded wavelength. A very large number of randomly-spaced heads would provide a set of vectors whose sum varied little with wavelength for any given setting of reverberation time, i.e. the frequency characteristic of the apparatus would be flat as regards these considerations. The greater the number of reproducing heads in the delay chain the more nearly can this ideal condition be approached. In the present machine, using eight reproducing heads, the maxima and minima from this cause are fairly numerous and closely spaced (as in the frequency response of a normal enclosure), and noticeable colorations do not arise unless two or more pairs of heads near the beginning of the delay chain, where the reproduced levels are high, are reproducing in phase at the same frequency. This would occur, for example, if the distances between the heads H and I and I and J (Fig. 8) were related in the ratio 1 : n, where n is a whole number. Then frequencies of recorded wavelength equal to the distance between H and I, or to sub-multiples of it, would be reproduced in phase by all three heads, and there would be a peak in the overall response of the apparatus at this fundamental frequency and its harmonics. A suitable arrangement of spacings between the heads is adopted to avoid this effect, due regard being paid to the flutter requirements previously described.

(14.2) Recording-Head Spacing and Coloration

Coloration also occurs when the recording heads, R_1 and R_2, are recording in phase on the medium, i.e. when the signals on the two tracks are in phase as they arrive at the first and subsequent reproducing heads. If the coils of the heads, R_1 and R_2, are fed in phase (i.e. in series-aiding) the distance between R_1 and R_2 must represent an integral number of wavelengths in order to produce coloration. If the heads R_1 and R_2 are fed in anti-phase (i.e. with their windings reversed) the distance R_1–R_2 must equal an odd number of half-wavelengths to produce coloration. In the present apparatus R_1 and R_2 are fed in anti-phase, and the distance R_1–R_2 is covered in 19·3 millisec. Therefore the fundamental coloration frequency of this arrangement is given by

$$f = \frac{10^3}{2 \times 19{\cdot}3} = 25{\cdot}9 \text{ cycles per second.}$$

Further frequencies of coloration will arise when R_1–$R_2 = 3\lambda/2$, $5\lambda/2$, $7\lambda/2$, etc., i.e. at 77·7 c/s, 129·5 c/s, 181·3 c/s, etc. Minima will also occur in the frequency response from this cause, for when the recording-head spacing equals one wavelength (anti-phase feeding) the recorded waveforms on the two tracks are in anti-phase, so that the two currents induced in the reproducing heads will cancel. This particular source of coloration is, of course, avoided in single-track working, but then the disadvantage of increased flutter (which is more noticeable subjectively) must be accepted. In any event, the broad maxima of this effect are broken up by the many other variations in the frequency response, such as that due to the reproducing heads discussed in the previous paragraph, and that due to the position of the feedback head, which is now to be considered.

(14.3) Coloration due to Feedback

The method of feeding back signals from a reproducing head into the recording heads, to enable a long reverberation time to be produced from a short-time delay element, necessarily introduces yet another source of coloration at various frequencies. In Fig. 10, R_1 and R_2 represent two heads recording on two tracks, H is any reproducing head spanning both tracks, whilst

O represents the feedback head which is feeding back from the lower track only. In considering the origins of reinforcement it is convenient to imagine that a steady tone is fed into the apparatus, so that both recording heads will still be recording tone from the external source when the head O is feeding back the reproduced signal into them. Reinforcement then occurs when the feedback head returns a signal into R_1 and R_2 in phase with the signal being fed to them by the external source, i.e. when the distance O–R_1 is an integral multiple of the recorded wavelength. Thus maxima in the overall frequency response will arise when O–$R_1 = \lambda$, 2λ, 3λ, etc., whilst minima will occur when O–$R_1 = \lambda/2$, $3\lambda/2$, $5\lambda/2$, etc., when the signal from O is in anti-phase with the external signal entering the recording head. This case corresponds to the coloration produced in the tube system discussed in Part 1. The degree of coloration, or the amplitude of the undulations in the frequency response, depends on the level fed back, so that both become more marked at long reverberation times.

The cyclic rise and fall of level with frequency which arises from this cause will be superimposed on that due to the recording-head spacing discussed previously. It is obviously undesirable that maxima due to the two causes should coincide, and the position of the feedback head is therefore chosen to provide minima at the frequencies at which the disposition of the two recording heads provides maxima. This is achieved by making the distance O–R_1 in Fig. 10 an exact odd multiple of the distance R_1–R_2. Thus at frequencies of 25·9c/s, 77·7c/s, etc., where

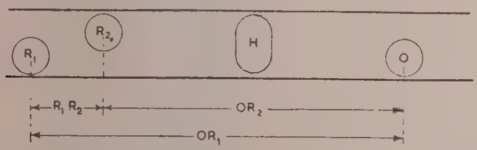


Fig. 10.—Distances governing occurrence of colorations.

maxima occur owing to recording-head spacing, the signal fed back to the two heads is in anti-phase to that being fed from the external circuit. In the present machine the distance O–R_1 is 13 times R_1–R_2, so that maxima (and minima) occur owing to feedback at intervals of just under four cycles, with minima falling at 25·9, 77·7c/s, etc., as required.

(15) SEPARATION BETWEEN THE HEADS AND THE MEDIUM

It has been shown that to avoid amplitude flutter in the artificial-reverberation system and to make conditions more realistic, the time spacing which must be arranged between the original sound and the first head "reflection," or between two adjacent reflections, must be as small as possible.

A given design of reproducing head determines the minimum possible distance between the gaps of adjacent heads, and the minimum time delay which is required must therefore be achieved by a suitable choice of recording speed. From considerations of simplicity, cheapness and sensitivity, conventional recording heads about 1in in diameter were chosen for the present apparatus. This dimension, in conjunction with the maximum time delay of about 30millisec permissible between adjacent reflections, dictated a recording speed of 100in/sec. Any loop or rotating surface of conventional size would have to be frequently renewed if recording and reproduction were carried out in contact with some 11 heads (such as are used in this system) at this speed. Consistency of performance and ease of maintenance thus necessitate out-of-contact working. Various factors not normally met with in conventional magnetic recording then become of importance. The effects of separating the head from the recording medium are most marked in the reproducing process, the response at any given wavelength being proportional to $\exp(-2\pi x/\lambda)$, where x is the separation of the recording medium from the pole tips of the reproducing head and λ is the recorded wavelength. This amounts approximately to a loss of 55dB of signal per wavelength of separation from the medium. At lower speeds such a loss would be very serious, but in the present apparatus, where the recording speed is high, the losses in the range of wavelengths which are of interest are acceptable. This range is determined, as noted in Part 1, by the absorption properties of a normal enclosure, which introduce a marked attenuation of sound frequencies above about 4kc/s when successive reflections take place. In the artificial-reverberation system, therefore, the output from the reproducing heads can begin to fall as the frequency rises above about 4kc/s. The loss at high frequencies which occurs with separation can, in fact, be utilized with convenience to provide the required characteristic.

In the case of the erasing and recording processes, separation from the medium means increasing the power fed to the recording and erasing heads to obtain the required field strength. Too large a separation involves very high power and can result in over-heating of the recording and erasing heads.

Another factor determining the separation chosen arises in the manufacturing process, where the limits of accuracy to which the disc circumference can be turned and ground set a minimum to the spacing. Excessive eccentricity of the disc in its bearing, apart from necessitating a corresponding separation, will introduce an appreciable amplitude modulation of the signal as the recording medium approaches and recedes from both the reproducing and the recording heads. With reasonable care in manufacture the eccentricity of the wheel can be kept fairly consistently below 1/5 000in, so that the separation distance could be of this order if required. In practice, a figure of 1/1 000in has been chosen, which reduces the possibility of damage to the magnetic coating in the setting-up process or as wear takes place in the bearing, and is satisfactory with respect to the other considerations discussed.

(16) FURTHER FACTORS AFFECTING THE OVERALL FREQUENCY RESPONSE

Various factors which affect the overall frequency response of the apparatus have been examined in previous Sections, notably those concerned with separation and with resonance effects which arise from the use of the two recording heads, the many reproducing heads and the feedback system. In addition, the normal factors which determine the overall frequency response in any magnetic recording system are present and in some cases in an exaggerated form not met with in conventional recording systems. These exaggerated effects arise from the high recording speed, and hence the long wavelengths, which are employed. For example, when the wavelengths become very large compared with the size of the reproducing head an appreciable proportion of the flux emanating from the recording medium does not proceed through the core but flows around the back of the head and so does not link with the coils in such a way as to produce a useful voltage. In the present apparatus the diameter of the reproducing-head core is about 1in and the wavelength at 30c/s is some 3in, so that the effect is becoming important near the bottom of the frequency response, resulting in a loss of signal/noise ratio. More important, however, are the interference effects connected with the structure of the reproducing head and screening box. The interference effect which occurs in the normal gap of the reproducing head is well known. This results in zero output from the reproducing head when the effective gap length d is an even multiple of half the wavelength being reproduced, and maximum output when the effective gap

length is an odd multiple of half the wavelength being reproduced, for the factor, sin $(\pi d/\lambda)$, which determines (approximately) the output at shorter wavelengths, is zero when $d = \lambda$, 2λ, etc., and unity when $d = \lambda/2$, $3\lambda/2$, etc. In the long-wavelength range, which occurs at the lower frequencies, this particular gap (or discontinuity in the flux path) ceases to have much significance, however, and larger secondary gaps assume importance. These secondary gaps are formed by other symmetrical discontinuities in the path of the flux entering and leaving the head. Fig. 11 shows the plan view of a conventional reproducing head con-

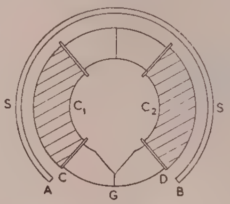


Fig. 11.—Plan view of conventional reproducing head and screening cover.

tained within its Mumetal screening box SS. The Mumetal screen is open between A and B in order that the gap G may be brought into proximity with the recording medium. When the recorded wavelength is much larger than the width G, the spread of flux from the tape is considerable so that it enters the core of the head not only through the gap G but also through the areas between C and G and between G and D. A second effective gap, defined by the ends of the coils C and D, can then become important, and the opening AB of the screening box may provide yet a third effective gap. The frequency response which results cannot be equalized by conventional means without serious reduction of signal/noise ratio. Now the configuration of the delay unit is such that AB must be fairly wide to allow the normal gap G, which is the operative gap at short wavelengths, to be brought into the required proximity to the magnetic track. If the dimensions CD and AB then become roughly equal, the two secondary gaps combine to create very marked undulations in the response. Conversely it has been found possible to reduce the undulations to acceptable proportions by the simple expedient of making the lengths CD and AB in the ratio of 2 : 3, when, at the lowest frequencies, the minima created by one reduce the maxima created by the other, and for the present practical purposes, satisfactory equalization is achieved. It is found that the further widening of the gap AB to conform to this requirement does not, in the present design, seriously increase unwanted pick-up of hum or other signals.

Part 3. THE REDUCTION OF FLUTTER EFFECTS IN ARTIFICIAL-REVERBERATION SYSTEMS

(17) INTRODUCTION

In Part 1 it was shown that an impulsive sound passed through an artificial-reverberation system, involving transmission in a single dimension, will be heard as a flutter when the successive reflections generated in the system are separately audible. This effect is reduced in the design of the magnetic system by the use, already described, of double-track recording and by careful attention to head spacing. Nevertheless, it still remains an obstacle to the use of the equipment in connection with some types of broadcast programme.

This Part of the paper deals more fully with the flutter effects and describes auxiliary equipment by means of which they are virtually eliminated by the interposition of large numbers of reflections between those generated in the artificial-reverberation equipment.

(18) METHODS OF REDUCTION OF FLUTTER

(18.1) Survey of Possible Methods

It will be appreciated from the previous discussion that the effects and severity of a flutter arising in an artificial-reverberation system depend upon the spacing of the separate reflections. It has also been noted that, in the magnetic system, the time interval between reflections is of the order of 15 millisec, and at this repetition rate it is often possible to hear the successive reflections of an impulsive sound.

The problem therefore becomes that of generating a large number of additional reflections, irregularly spaced, to fill the intervals between the reflections from successive reproducing heads. This not only follows from the discussion above but is closely analogous to the problem of "flutter echo" in room acoustics. It has been known for a long time that flutter echo will occur in any room having one pair of opposite surfaces sufficiently reflecting. Recent investigations in the B.B.C. Research Department have shown that it is the ratio of the reflection coefficient of one pair of walls to that of the other pair which is significant; a reverberation room having all its surfaces highly reflecting is free from flutter echo, whereas a highly-damped talks studio may cause flutter if the absorption coefficient of one opposite pair of surfaces is lower than about 0·65 times the mean of the other pair. A cure may be effected in this case either by reducing the absorption coefficient of the more highly absorbing walls, or by increasing that of the reflecting pair, until equal reflections are obtained from all three directions.

It follows that the most promising method of introducing the required additional echoes is to make use of actual reflections from the inside walls of a room or box, thus adding the effect of two more dimensions. The requirement that there should be numerous reflections within a time limit of 15 millisec implies that the room should not be very large, but too small a space is also undesirable owing to the probability of severe colorations.

(18.2) Reduction by Interposition of Room Reflections

A preliminary design experiment made use of rooms of 1 000–2 000 ft^3 in volume, the original sounds being radiated into the room by means of a high-quality loudspeaker and picked up by a moving-coil microphone in one corner of the room. The output of the microphone, consisting of the original signal and a series of reflections from the six walls, was then passed to the magnetic reverberation equipment. A small reverberation room with a mean reverberation time of 3 sec and an experimental talks studio of 0·35 sec reverberation time were found to be equally effective in suppressing the flutter, but only the talks studio was sufficiently free from colorations. From previous study of coloration in talks studios, it was judged inadvisable to reduce the room dimensions or to increase the reverberation time to any great extent. The addition of a normal-sized room to the artificial-reverberation equipment, however, would clearly offset many of its advantages, and indeed, for many purposes render it superfluous. Other means of providing short-term reflections within a small space of time without adding serious colorations were therefore sought.

(18.3) Heavy-Gas-Room Model

The minimum permissible dimensions of the room, being dependent on the wavelength of the sound, are reduced if the air is replaced by a medium in which sound travels more slowly. The only media having sound velocities substantially lower than that in air are gases, notably radon (0·38 of that in air), chlorine

(0·64), krypton (0·59), xenon (0·47), sulphur hexafluoride (0·45) and a few organic gases (down to 0·49). In no instance was the reduction in velocity, and therefore room dimensions, sufficient to justify further consideration of this method.

(18.4) Flat Model

Two other methods were accordingly tried, both of which may be regarded as practical solutions to the problem. In the first, the room was replaced by a flat rectangular cavity in which two of the dimensions were comparable with living-room dimensions, but the third was only 10cm. The wanted reflections were provided by the four narrow walls of the cavity, cross-reflections between the two large walls having little effect. This form of construction requires less space than a room with three comparable dimensions; it could be built in the form of a cavity wall between two rooms.

The reverberation time of the laboratory model was adjusted to a value of 0·3sec at all frequencies within the range 100–5 000c/s, and the positions of the loudspeaker and microphone were chosen by listening tests to give minimum coloration. The performance was tested thoroughly using pulses of tone, impulsive noises, octave bands of white noise, speech and musical programmes. It was found to be almost entirely successful in eliminating flutter effects, and was fairly free from serious colorations.

(18.5) The Ultrasonic Reverberation Tank

The second practical method[12] to be developed was to replace the echo room by a small tank filled with water, and to transmit the programme through the water in the form of a modulated carrier. By this means the wavelength of the signal in water could be reduced to a small fraction of the dimensions of the tank, and simple standing-wave systems avoided. A full description of the equipment, which is now undergoing trials in the Sound and Television Services in conjunction with the artificial-reverberation apparatus, follows in the next Section. Its performance is better than that of the flat model, and it has the additional advantages of transportability and adaptability.

(19) THE ULTRASONIC REVERBERATION SYSTEM

(19.1) General Description

Fig. 12 is a block schematic, and Fig. 13 a photograph, of the first complete equipment. The original signal is separated into two parts, one going through a 250c/s low-pass filter to the output mixer, the other passing through a complementary high-pass filter to the ultrasonic system before reaching the mixer. The reason for this will be made clear. After modulation of the carrier signal by the programme, the upper sideband only is selected and fed to a barium-titanate transducer dipping into water contained in a glass tank, through which it is transmitted as ultrasonic wave motion; a similar transducer reconverts this, together with numerous reflections from the sides of the tank, into electrical signals. Finally, the output of the tank is demodulated by the receiver and recombined by the mixer with the low-frequency part of the original signal, before passing to the artificial-reverberation apparatus. If desired, the ultrasonic system may be introduced at the output of the reverberation apparatus instead of the input as here described; the effect is the same, and the position may be decided solely from the standpoint of engineering convenience.

(19.2) The Acoustics of the Ultrasonic Tank

The main design features of the tank were determined by the requirement that the whole reverberation system should be transportable, and preferably that it should be accommodated in an ordinary apparatus bay. For the system to be reasonably free from the worst standing-wave effects, the greatest wavelength of the sound in the tank should be small in comparison with the smallest dimension. If the tank were filled with air, the lowest permissible frequency would be of the order of several thousand cycles per second. A carrier system is clearly indicated as the only means of making a substantial reduction in wavelength, and in order to avoid acoustic interference from its surroundings, including the moving parts of the reverberation apparatus itself, the transmitted band should be outside the audio-frequency range, say above 15kc/s. At this frequency the attenuation of sound in air, which increases as the square of the frequency, already limits the possible reverberation time to approximately 0·4sec, and at the upper limit of the band the highest possible time would be 0·2sec. Any absorption of sound by the walls would reduce the reverberation time to lower figures. Furthermore, in order to obtain a flat response over the required band of frequencies, it would be necessary to use piezoelectric transducers without appreciable tuning; the insertion loss would therefore be large, and it would be difficult to achieve a good signal/noise ratio. For these reasons, it was decided to

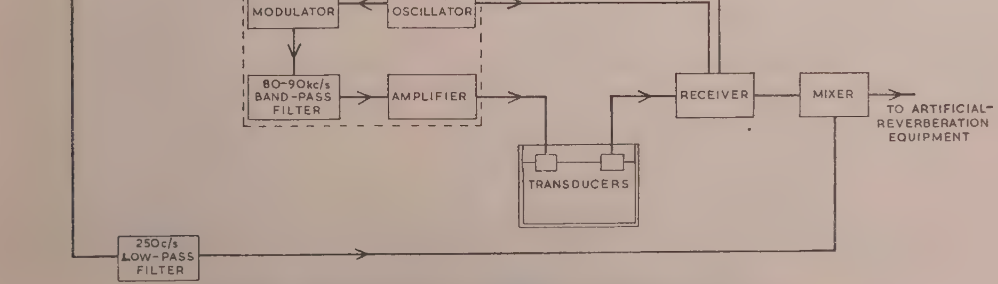


Fig. 12.—Block schematic of ultrasonic reverberation tank chain.

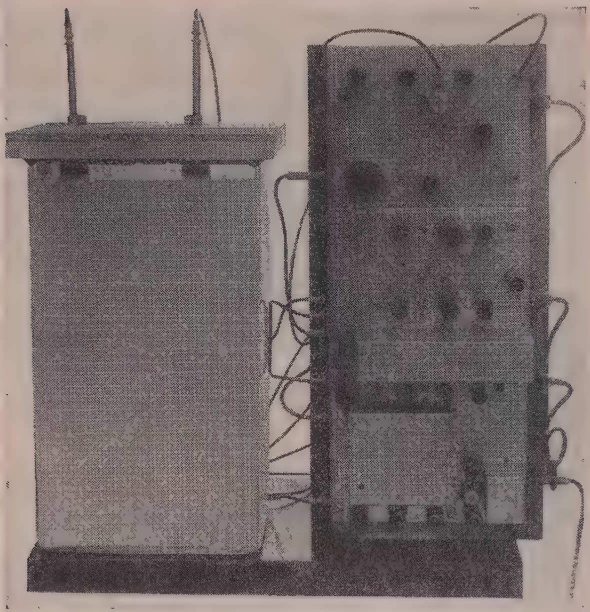


Fig. 13.—Ultrasonic reverberation tank apparatus.

use water as the transmission medium in the tank. The attenuation loss in water is only one-eightieth of that of sound of the same wavelength in air, and by good design the insertion loss of the two transducers with a water path between them may be made very small. The velocity of sound in water is, however, about 4·3 times that in air, and it is thus necessary to increase the frequency in the same ratio to obtain the same wavelength.

The tank at present in use is a rectangular glass accumulator vessel, with a base measuring 30cm × 20cm and a height of 50cm to the surface of the water. A carrier frequency of 80kc/s was chosen, thus permitting the use of tuned crystal transducers of convenient size.

(19.3) Standing-Wave Systems and Reverberation in the Tank

At this frequency the wavelength of sound in water is approximately 2cm, giving ten wavelengths across the smallest dimension of the tank. The total number of possible normal modes in a 500c/s bandwidth on either side of this frequency is 320—a number of the same order as that in the band 0–500c/s in a room and sufficiently large to ensure that they would be too close to be separately distinguishable if all were of comparable significance. Mayo's work,[13] however, shows that the harmonics of the axial modes will be very much more important than any individual tangential or oblique modes, and in an ordinary room they are the only audibly significant modes. The remaining modes in a room do, however, serve a useful purpose in acquiring and dissipating some of the reverberant energy which would otherwise be concentrated in the axial modes, with the result that the colorations in a room are less severe than those heard from a combination of three delay systems as described in Section 7.2.

With the ultrasonic tank, the axial modes may be shown to be relatively more important than with the room, but they are more widely spaced within the frequency band, and it is found that the colorations may be made small by choice of transducer positions.

Reflection of the sound at the outside surface of the tank is almost complete, owing to the great differences between the characteristic resistance of the glass or water and that of the air outside, and since the attenuation in water is so small, the present reverberation time of 0·3sec was very easily attained. By attention to detail, much higher values are possible. Materials other than glass would be equally satisfactory for the container; a galvanized iron cistern sprayed with a high-gloss paint was used for the early experiments giving reverberation times up to 0·75sec, and a non-corroding thin sheet metal will be used in future versions.

(19.4) Distortion due to Double-Sideband Modulation

In the first experimental system, the 80kc/s carrier was amplitude-modulated by the a.f. signal, giving double sidebands (74–86kc/s). This resulted in a very serious distortion, similar to the selective fading heard on long-distance radio transmissions, because the two sidebands corresponding to any given frequency reached the receiving transducer with different phases and amplitudes. It was therefore decided to change to single-sideband modulation, and the adoption of this system as described resulted in acceptable transmitted quality.

Experiments were also made on the reduction of coloration by the use of rectangular projections on the walls of the tank. The results were promising, and diffusers of this type will be incorporated in the sheet-metal tanks now being designed.

(19.5) The Transmission Chain

The required frequency range of the ultrasonic tank is limited at both ends. The upper limit need only be high enough to correspond to the highest frequency transmitted by the reverberation apparatus with which it is used. The lower limit is chosen from a consideration of the spacing between the reproducing heads of the magnetic-recording apparatus. Flutter is heard only when the duration of a sound is short in comparison with the time spacing between the heads, and it follows that the lower-frequency spectral components of the sound, which have periodic times comparable with the spacing of the heads, do not contribute to the flutter. On the other hand, they are affected by any remaining colorations due to the ultrasonic tank, and

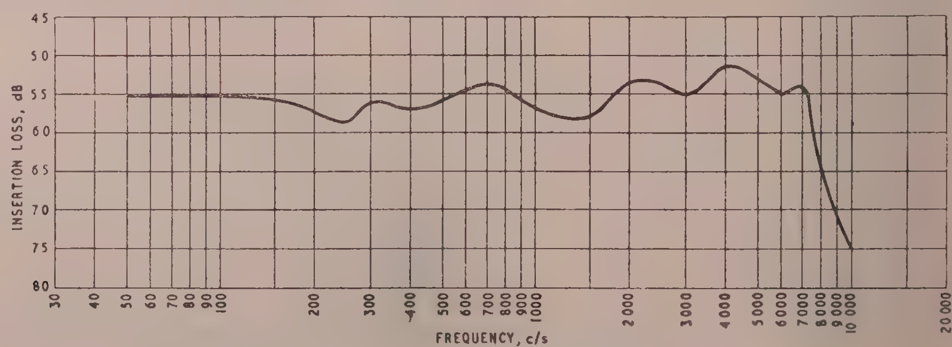


Fig. 14.—Frequency characteristic of ultrasonic system.

it is the low-frequency colorations which are invariably the most obvious and persistent. The low-frequency components of the incoming programme are therefore separated from the rest by the low-pass and high-pass filters shown in Fig. 12, the low-frequency components being transmitted through a side chain and recombined with the high-frequency components after they have been through the flutter-suppressing equipment.

The output of the 250c/s high-pass filter is fed to a balanced ring-modulator supplied with 15kc/s tone from an oscillator in the same chassis. The modulated output consists of two sidebands, the upper sideband, 15·25–25kc/s, and the lower one, 5–14·75kc/s, together with any remaining carrier which has not been completely balanced out. The 15·25–25 kc/s band-pass filter selects the upper sideband, which is passed to the second modulator where it modulates a 65 kc/s oscillation generated in the same chassis, giving sidebands at 80·25–90 kc/s and 40–49·75 kc/s. The latter is removed by the 80–90 kc/s band-pass filter.

The two-stage modulation was adopted in the prototype equipment because single-stage modulation would, of course, have given a lower sideband extending up to 79·75kc/s. To remove the 80kc/s carrier and the sideband from 79·75kc/s downwards without affecting the lower end of the upper sideband would require the development of a special quartz-crystal electromechanical filter giving an extremely sharp cut-off. It was possible, however, to design an electrical filter using Ferroxcube inductance cores to perform the same operation, provided that the carrier frequency was not higher than about 15kc/s, and the double modulation system was therefore used as described. The limitation of the frequency band to 250c/s at the lower end had the additional advantage of simplifying the design of this filter.

The output from the receiving transducer is passed to the receiver (Fig. 12), to which is also fed in correct phase an 80kc/s carrier derived from the 15 and 65kc/s oscillators, and the resulting a.f. signal passes to the output mixer. The frequency characteristic of the whole chain is shown in Fig. 14.

(19.6) The Transducers

The transducers consist of tubes of barium titanate, each having inside and outside diameters of 1·905 and 2·540cm. Their lengths, approximately 2·3cm, were adjusted by grinding to give staggered resonance frequencies of 84 and 86kc/s respectively, thus flattening the overall response of the system. The inner and outer curved surfaces of the crystals are silvered, and contact is made by spring pressure and soldering respectively at points on the transverse plane of symmetry, as shown in Fig. 15. A radial field thus applied causes changes of length by the Poisson

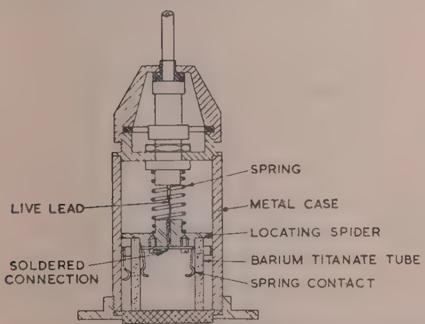


Fig. 15.—Section through barium-titanate transducer.

effect. The tube is mounted in a watertight container, one end being held by light spring pressure against the diaphragm through which the vibrations are transmitted. These transducers are very well matched to the water, the combined insertion loss being of the order of 10dB.

(20) GENERAL CONCLUSION

The discussion in the paper has underlined many of the technical difficulties which must be overcome in order to provide an artificial-reverberation system, other than a reverberation room, giving realistic quality. When these technical difficulties can be surmounted a choice is available between a reverberation room and some form of artificial-reverberation system involving one or more delay channels. If long reverberation times are required the reverberant room itself must be of fairly large dimensions, and there is therefore great economic advantage to be gained by installing a comparatively small delay system and using the room space for other purposes. This economic advantage may be very great where broadcasting premises are in densely populated areas such as Central London, where rateable values are high and the possibilities of expansion are limited. If a choice is made in favour of an artificial-reverberation system using delay channels, consideration must be given to the type of delay system which is to be adopted. In Part 1 a description has been given of a delay mechanism consisting of an acoustic tube. A performance adequate for some types of programme can be obtained with this system whilst still maintaining the bulk and capital cost of the apparatus within reasonable limits. A completely general application of this form would, however, require an increase of the bulk of the acoustic tube, and of the complexity and cost of the apparatus associated with it, to a degree where the economic advantages of the artificial method would no longer be very marked. A rather smaller and more convenient form of delay mechanism is required, and, as shown, a magnetic recording system can be designed to fulfil this function. With this system, comparatively long reverberation times may be obtained with a delay unit of small bulk. The unit may be designed without unreasonable complications to reduce the fundamental amplitude flutter to proportions which make it acceptable on a very wide variety of programmes. There are still some limitations, however, and the machine described in Part 2 has not proved entirely suitable when used alone in drama productions, when much reverberation has to be added to the output from a fairly dead sound broadcasting studio. To meet this situation, however, the ultrasonic reverberation system has been developed to work in conjunction with the artificial-reverberation machine. The reduction of residual amplitude flutter by this means seems to hold great promise. Moreover, the ultrasonic tank is of small dimensions, and together with its electronic equipment, it can be mounted in a common cabinet with the magnetic recording system. This is possible if the height of the 5ft cabinet in which the apparatus is housed is increased by only some 2ft.

It would seem, therefore, that the combination of these two equipments may find increasing application in the broadcasting service.

Operational experience with the magnetic-recording reverberation machine alone is now quite extensive in the Television Service, and it is used regularly on all types of programme except when "cathedral" or exceptionally delayed echo effects are required. Here the amplitude flutter becomes noticeable, but information is not yet available as to the influence which the ultrasonic unit may have on this restriction. In sound broadcasting less operational experience is available, but the machine has proved successful on musical programmes. Of some interest is its use to increase the brilliance or reverberation in certain sections (e.g. the strings) of light orchestras to give novel effects of orchestration.

The machine is also employed in other roles not concerned with immediate transmission. In feature and survey programmes, for example, composite dubbings are frequently prepared using direct recordings from various sources in which the original reverbera-

tion conditions were widely different. Comment or announcements may also be required to be inserted between the various excerpts from the originals. Experience has shown that a wide variety of reverberation conditions occurring in the course of such a composite recording may have a disturbing and unrealistic effect. The reverberation machine is then useful in equalizing the various reverberation conditions in the make up of the programme.

(21) ACKNOWLEDGMENTS

The authors wish to acknowledge the contributions made by other members of the B.B.C. Research Department to the development described in this paper, and in particular, the work of Messrs. E. S. Burnett, H. D. Harwood, C. Henocq, W. I. Manson, R. A. Ticehurst and F. L. Ward.

The paper is presented by kind permission of the Chief Engineer of the British Broadcasting Corporation.

(22) REFERENCES

(1) RETTINGER, M.: "Reverberation Chambers for Re-Recording," *Journal of the Society of Motion Picture Engineers*, 1945, **45**, p. 350.

(2) WOLF, S. K.: "The Synthetic Production and Control of Acoustic Phenomena by a Magnetic Recording System," *Proceedings of the Institute of Radio Engineers*, 1941, **29**, p. 365.

(3) HILLIARD, J. K.: "The Variable-Density Film Recording System used at M.G.M. Studios," *Journal of the Society of Motion Picture Engineers*, 1943, **40**, p. 143.

(4) OLSON, H. F.: U.S. Patent No. 2493638. January, 1950.

(5) CURRAN, G. W.: "An Artificial Reverberation System," *Audio Engineering*, 1948, **32**, No. 5, p. 13.

(6) GOLDSMITH, A. N.: "U.S. Patent No. 2105318. January, 1938.

(7) BERTH-JONES, E. W.: Patent No. 636223. April, 1950.

(8) HOUSTON, R. S.: "Echo Effects with a Tape Recorder," *Audio Engineering*, 1952, **36**, No. 5, p. 52.

(9) GOLDMARK, P. C., and HENDRICKS, P. S.: "Synthetic Reverberation," *Proceedings of the Institute of Radio Engineers*, 1939, **27**, p. 747.

(10) AXON, P. E., BURNETT, E. S., HENOCQ, C., and MERRY, I. W.: Patent Specification No. 29270, November, 1952.

(11) STOTT, A.: Patent Specification No. 22871, September, 1952.

(12) GILFORD, C. L. S., and TICEHURST, R. A.: Patent Specification No. 13165, May, 1954.

(13) MAYO, C. G.: "Standing Wave Patterns in Studio Acoustics," *Acustica*, 1952, **2**, p. 49.

DISCUSSION BEFORE THE RADIO SECTION, 9TH MARCH, 1955

Dr. G. F. Dutton: I am very glad to see a paper dealing with this very old established problem of artificial reverberation. Since the introduction of the microphone into the science and practice of sound recording and transmission, the listening public has voiced its opinion on the amount and character of the reverberation. The lay public can discuss it at length, because it is a matter of individual taste and requires no special knowledge of music or artistic training to argue on the point. It is simply a question of like or dislike, and because of this, opinion may be somewhat swayed by the prevailing fashion.

The authors have laid down certain rules which give the preferred characteristic, but it might be worth while attempting to state the reasons for the desire on the part of the listener for reverberation. The listener prefers some reverberation because he likes to hear some continuity of sound and to have some response from the studio. In general, however, the golden rule is that reverberation should be heard but not noticed.

There is no doubt that a better performance can be obtained on orchestral recordings or transmissions of recordings in an acoustically correct environment. Under these conditions the orchestra is encouraged to give of its best. Artificial reverberation could serve a very useful purpose, however, in correcting studio reverberation faults and adding effects for dramatic performances.

The reverberation-chamber or echo-room method of studio correction encounters many difficult problems, as the authors have clearly stated. All delay methods introduce transducers in the system, with their attendant problems. An echo room supplies one reproduction from the loudspeaker, whereas the acoustic-delay tube and the magnetic-delay systems supply many cycles of recording and reproduction. For instance, if the delay time is 30 millisec and the reverberation time is 2 sec, the number of re-recordings will be over one hundred. Fortunately the level of these recordings is sufficiently low so that distortion can be made negligible.

There is, I think, a tendency in the paper to underestimate the importance of the higher frequencies in artificial reverberation required for studio correction. For this purpose, the frequency range should extend to 10 kc/s, and it may well be that the correction required forms a characteristic rising with frequency.

The effect of wow and flutter on the magnetic-delay system is not treated in the paper, but I imagine that it is not a serious problem.

Since the paper was published and I have seen in it the statement that the equipment was being used continuously on television programmes, I have listened to several programmes, the character being such that I would have expected the producers to use artificial reverberation. There was one vocal item in the "Quite Contrary" programme which I think was spoilt by reverberation of a drain-pipe character. The long period of this reverberation seemed to bear no relationship to the sound one would naturally expect from a close-up view of the singer. On the other hand, I listened to the Vienna Choir and could not detect any added reverberation. I personally would have preferred a long singing quality in the reverberation on this item, but perhaps the equipment was not being used in this particular programme.

The method of magnetic delay described by the authors should be an invaluable research tool for investigating studio acoustics, particularly on the psychological side.

Mr. P. Bate: In the field of television I deal with a number of orchestral programmes for an audience who subconsciously desire—indeed demand—something approaching concert-hall acoustics, or what they imagine concert-hall acoustics to be. For economic reasons and so on, we are faced with the prospect of using regularly for that type of programme a fundamentally unsuitable studio. Therefore we must do our best to improve it in relation to what we feel to be right and what the audience demands. This apparatus represents a great step forward in that direction.

The authors state that the assessment of sound quality is ultimately a matter for the ear only, i.e. it is a subjective matter. Is there any comparative method of measuring or tabulating a consensus of opinion among users and audiences such as those I try to serve? It would be of great value if something in reasonably compact and usable terms could be devised—or indeed has been devised. Would it not be equally useful as a guide to

development engineers and those concerned with meeting our practical problems?

Dr. R. C. G. Williams: The average popular recording consists of orchestration, vocalist, sound effects and artificial reverberation, and the art of the producer is to obtain the mix of these ingredients which will produce the most pleasing overall result. Most recording companies still use echo chambers to obtain their reverberation effects. As the authors have indicated, these are inherently inflexible, and we are indebted to the authors for an excellent piece of analytical research into reverberation as well as developing apparatus which provides such a wide degree of control and adjustment. Their work makes available to the record producer a new tool of general application which will enable him to use the degree of reverberation of his choice, either for realism or for effect.

An important distinction between reverberation applied to broadcasting and to recording now becomes apparent. For broadcasting the performance is usually "live," and reverberation if used must be superimposed during transmission, but for recording it can, and in fact normally is, added afterwards between the original tape recording and the cutting of the acetate master record, in the same manner as is used for other background effects. In addition, the amount of reverberation and the frequency range over which it is applied can be worked out without undue haste on a "trial-and-error" basis. This rather weights the scales in favour of echo chambers for recording, since they can be built or halls can be hired in a remote location or at a time when cost is at a minimum.

Would the authors comment on this point and perhaps give something of their experience when starting up this work and when they were investigating the echo chambers then in use? In particular, does their organization use artificial echo effects on recorded programmes as a subsequent addition, in the same way as is done in the gramophone record industry?

I remember the first occasion on which I noticed artificial reverberation in a television broadcast of a well-known woman vocalist. It was perhaps particularly obvious to me, since I had become accustomed to the difference between performers' voices as heard over the air and as normally recorded. Whatever may be our personal views on the professional ethics of a deliberately unnatural recording, it is this "treated" voice which the public associates with the performer through records, and I would suggest that some co-operation in this respect between the recording industry and the broadcasting authorities might be of value.

Mr. W. H. Livy: In my opinion, with this system the reverberated sound is very much more noisy, and the background noise is considerably higher than that of the "dry" sound. What degree of noise increase is generated by the system?

Furthermore, the sound is somewhat distorted, particularly in the higher frequencies of the reverberated sound. What increase in distortion is produced by the system? These two points can be grouped together by giving a figure for the signal/noise ratio of the reverberation channel for a specified distortion content in the output.

Mr. G. Corran: Is the performance diagram based on practical tests or theory? If the latter is the case, could the authors state how closely the figures agree with observed results?

Reverberation rooms usually conform to conventional shapes, for obvious reasons. Have the authors any information on the use of a spherical echo chamber?

The separation of the recording and reproducing heads from the magnetic surface in the magnetic recording delay unit is very critical. Has any difficulty been experienced in maintaining the correct clearances on these heads?

Could the authors give a figure for the signal/noise ratio obtainable on the magnetic delay unit and on the ultrasonic reverberation equipment?

Mr. F. Williams: The system has been used on television broadcasts for some time. For musical programmes on television it has a great application, since many television studios are necessarily rather dead acoustically. They must be, since, in order to keep the microphone out of view on speech, it has to be used at a greater distance than in corresponding conditions in sound studios. If the studio were reverberant, the speech would be more reverberant than would be acceptable.

In sound broadcasting, on the other hand, one tries to arrange for music to be played in a studio which has the best reverberation for that size of orchestra. Therefore, the application of artificial reverberation on music in purely sound broadcasting is not the same as in television. However, there is an application for special effects on music when one wants to bring out a particular section of the orchestra or to get over special effects on solo instruments.

For music the apparatus is found quite acceptable. For drama, however, in our experience the prototype model is not yet perfect. There is still, in spite of the addition of the ultrasonic tank, a sensation of flutter with pistol shots and other such impulsive sounds. Do the authors anticipate that development can be carried to the point at which the machine would in fact be completely acceptable for that type of sound required in dramatic productions?

How does the cost of the equipment compare with the normal reverberation chamber? The authors have drawn attention to the problem of economics, and it does apply particularly in built-up areas like London, where the cost of land is high. Even so, these reverberation chambers can be located in basement areas which are not so valuable. They are simple and there is nothing really to go wrong with them. A really well-proportioned room can give an acceptable reverberation on almost any type of programme.

Such a room could have dimensions of, say, $20 \times 15 \times 12$ ft. That type of room can be made to give, for dramatic productions and also for many types of music, a really good reverberative sound which is preferred by many drama producers to the type of sound which they get from this reverberation equipment. The sound from the equipment is rather synthetic. Subjectively, the reverberation time does not seem to be as long as it is theoretically calibrated to be. It does not seem to be quite so "full-blooded."

One reason for this may be that the apparatus has been designed for a reverberation characteristic which is truly exponential. Would it be better if the decay characteristic were such that the level of energy was "held up" rather above that exponential curve and then dropped somewhat more suddenly, and by that means could we get a rather more "full-blooded" effect?

THE AUTHORS' REPLY TO THE ABOVE DISCUSSION

Dr. P. E. Axon, and **Messrs. C. L. S. Gilford** and **D. E. L. Shorter** (*in reply*): We would agree with Mr. F. Williams that the ideal solution to the problem of providing optimum reverberation is to construct studios which have in themselves the required characteristics. We are, however, unable to support his optimistic appraisal of the use of reverberation rooms in meeting the situation when a suitable studio is not available. It was, in fact, a realization of the disadvantages of this form of artificial reverberation which led to the developments described in the paper, and which has stimulated similar research

on the Continent and in the United States. Mr. Williams implies that a correct proportion of dimensions of the reverberation rooms is sufficient for a satisfactory effect. The principal shortcomings of reverberation rooms arise from their small dimensions in relation to the size of enclosure they are intended to simulate. This results in too wide a spacing of the mode frequencies and an absence of long-path reflections. A highly reverberant room of the suggested dimensions of 20 ft × 15 ft × 12 ft would, in fact, sound more like a washroom than a large studio. The apparatus would not be improved by departing from an exponential decay, since this is one of the desirable characteristics of studios and reverberation rooms alike. Moreover, the general trend of the decay in any simple linear system employing feedback must be exponential, for the signals circulating in the delay loop suffer equal amounts of attenuation at regular intervals of time (Section 6). A combination of several such systems having different reverberation times could provide a rate of attenuation greater at the start than at the end of the decay, but the converse effect suggested could not thus be achieved. There is no doubt that apparatus could be developed in which the flutter effect on transient sounds was even less noticeable than in the present combination of the magnetic-recording delay unit and the ultrasonic tank. The most promising line of development is a fairly large increase in the size of the latter. Comparisons of cost of the present equipment and of reverberation chambers are difficult. It is impossible to generalize about the value of rates, rents and other overhead charges without knowing the situation and functions of the building in which the room is located, but in most large cities even basement premises are at a premium. The structural work required to produce a sound-insulated reverberation room in an existing, or proposed, building is very expensive since it is necessary to provide very thick or double walls and protection against structure-borne sound from other parts of the building. Such comparisons as we have been able to make have shown that the first cost of artificial-reverberation apparatus of the type described is much less.

A spherical reverberation chamber, as suggested by Mr. Corran, has been used in the film industry and is reputed, rather surprisingly, to have been satisfactory. The lower modes of a spherical space are widely separated and one would expect them to be very obtrusive.

In reply to Mr. Bate, it is not easy to establish a consensus of opinion on concert-hall acoustics owing to the variations of judgment encountered. However, some systematic investigations have been made recently,* and good agreement appears to exist regarding the optimum reverberation time for classical orchestral music.

* Somerville, T.: "Subjective Appreciation of Concert Halls," *B.B.C. Quarterly*, 1953, **8**, p. 125.
Kuhl, W.: "Nachhallzeit Grosser Musikstudios," *Acustica*, 1954, **4**, p. 618.

Dr. Dutton suggests that it is incorrect to allow the frequency characteristic of the reverberation apparatus to fall at above 5 kc/s. However, in actual large auditoria the reverberation characteristic falls sharply above this frequency owing to the high attenuation of high-frequency sound in air. It is possible to raise the higher-frequency response of the apparatus, but various design figures, chosen in relation to the natural characteristic, would have to be altered if any radical departure were required. It may be noted, however, that those early reflections in an auditorium which reach the listener with little high-frequency loss could be simulated, if desired, by an additional delay system having no feedback but transmitting the full frequency range. Wow and flutter are not a serious problem in the magnetic-delay element, for the mechanical system is fairly simple and the recording speed is high.

The effect of repeated recordings is not so serious as Dr. Dutton's example would suggest. His figure of 30 millisec for a 2 sec reverberation time represents a condition which is unlikely to obtain in any workable arrangement (see Section 6.1), since it calls for a loop loss of only 0·9 dB. In the magnetic system described, the overall delay is 250 millisec, and the corresponding number of re-recordings would be less than one-eighth of that suggested.

We must correct Dr. Dutton in attributing to us the statement that the equipment was being used "continuously." We have stated that it is used "regularly," i.e. on all current programmes in which the producers consider it necessary and suitable. As regards Dr. Dutton's complaint that a close-up view of a singer had not a natural sound associated with it, we can only disclaim responsibility and state that it has never seemed to us to be a prime aim of singers in the modern idiom to make themselves sound natural.

With reference to the increased noise and distortion which Mr. Livy claims to detect, the conditions under which these quantities are measured (or heard) must be carefully specified. In practice, the values must depend on the reverberation-time setting and the ratio of reverberant to direct sound from the studio. The reverberation chain can then only contribute a limited percentage to the final distortion and noise content of the output. Figures for comparison with normal recording equipment may be provided if the output of one reproducing head of the delay element is investigated. Under these conditions the total harmonic distortion of a 1 kc/s tone fed to the machine at peak level is less than 2% and the signal/noise ratio about 50 dB. A similar value of signal/noise ratio is achieved in the ultrasonic tank.

The apparatus has been used in a similar role to that described by Dr. R. C. G. Williams for the make-up of feature and survey programmes (Section 20).

req

621.395.625:534.862.3

Paper No. 1874 R
Sept. 1955

THE SUBJECTIVE DISCRIMINATION OF PITCH AND AMPLITUDE FLUCTUATIONS IN RECORDING SYSTEMS

By A. STOTT, M.A., A.Inst.P., and P. E. AXON, O.B.E., Ph.D., M.Sc., Associate Members.

(*The paper was first received* 31*st December*, 1954, *and in revised form* 7*th April*, 1955.)

SUMMARY

The paper is concerned with the measurement of subjective thresholds of the unwanted pitch and amplitude fluctuations which occur in the reproduced output of all types of sound recording systems. The pitch fluctuations are commonly known as "wow" or "flutter," and new equipment is described for generating controlled fluctuations of this type in musical or other programme signals. The threshold measurements thus made possible are more realistic than those hitherto available, which have been related only to frequency fluctuations in test tones. The generation and threshold measurement of controlled amplitude fluctuations are also described, since amplitude and frequency fluctuations commonly occur together. The results obtained enable frequency weighting characteristics to be defined which can be incorporated into instruments measuring the magnitude of the unwanted amplitude and frequency fluctuations in practical systems. It is shown that such instruments may then approximate to an ideal, subjectively weighted measuring device which, when threshold value is reached, will give the same indication irrespective of the type and frequency of the fluctuation measured.

(1) INTRODUCTION

The aural effect of the unwanted pitch fluctuations which occur in the reproduced output of all types of sound recording systems is well known. These fluctuations, which are usually described as "wow" or "flutter" according to the frequency range in which they occur, arise from various mechanical imperfections in the recording and reproducing systems. In disc recording the possible imperfections include small eccentricities in the driving system, eccentricity of the disc upon the turntable and lack of flatness in the disc surface. In magnetic recording the imperfections may consist of eccentricities in the driving system and of changes in tape speed due to frictional variations. All such factors result in periodic or random differences, or both, between the speeds of the recording medium in the recording and reproducing processes. In a recording machine a temporary departure from correct speed results in a change of recorded wavelength, which, on subsequent reproduction at correct speed, appears as a temporary change of pitch in the reproduced signal. In a reproducing machine a temporary speed change again results in a pitch fluctuation as the previously established waveforms move past the reproducing head at a rate different from that in the recording process. A characteristic of this process of frequency modulation is that all component frequencies of the recorded signal undergo an equal percentage modulation at the same time.

It is well known that these pitch fluctuations are, in practice, generally accompanied by amplitude fluctuations. In the work described tests have therefore been included on the subjective discrimination of amplitude fluctuations, to determine the influence of this factor on the final results. The information gained will be useful in the specification of permissible tolerances for pitch and amplitude fluctuations in recording equipment, a measure which is desirable for promoting general improvement in recording-system performance both in manufacture and in programme exchanges in broadcasting.

Experiments to determine the listener's sensitivity to pitch fluctuations necessitate the generation of wow and flutter under controlled conditions. In general, any type of pitch fluctuation may be introduced into a tone by appropriate alteration of some element in the oscillator circuit. No similar method is known by which an equal percentage frequency modulation of all the component frequencies of a complex programme signal can be effected, which is the practical characteristic previously described. A recording system, in fact, remains the practical method of meeting this requirement. However, in generating intentional pitch fluctuations by mechanical means, as opposed to electronic, a fundamental difficulty arises in eliminating the fluctuations inherent in the generator itself. For example, a good magnetic tape recorder might be used in which the driving shaft is made eccentric intentionally so that a periodic speed fluctuation is introduced into the recording. The record thus made might be reproduced on another good machine in which the driving shaft is true, with the result that a periodic pitch fluctuation at the frequency of the recorder driving shaft will appear in the output. On this periodic fluctuation, however, will be superimposed any other random and periodic fluctuations arising in the recording and reproducing machines. Experiments carried out in this manner could not therefore be said, without positive supporting evidence, to provide an accurate assessment of the subjective effect of the driving shaft fluctuation alone. In addition, of course, the provision of a wide range of fluctuations (in both magnitude and frequency), by this or similar means, would involve excessive labour. For these reasons the mechanical problems have hitherto forced investigators to use purely electronic methods, and this, in turn, has confined the investigations to tone tests. Recently, however, there has been notable development in the technique of recording (for storage purposes) on magnetic discs or drums rotating out of contact with the recording, reproducing and erase heads, and this has made possible the development of a pitch-fluctuation generator for use with complex programme material. The device has negligible inherent wow or flutter, but with it a wide range of controlled pitch fluctuations can be introduced into the programme signal.

(2) EXPERIMENTAL PROCEDURE

(2.1) General

The experiments were divided into four broad classes, in which known amplitude or frequency modulations were imposed on test material consisting of tones or programme of broadcasting type. The object of the experiments was to determine when modulation of the test signal was first noticeable; the subjective thresholds were measured under chosen conditions and were not necessarily associated with annoyance value. The effects of modulation may not always be displeasing from an aesthetic viewpoint. Although a prime object of the work was to establish

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Mr. Stott and Dr. Axon are with the British Broadcasting Corporation.

tolerances to modulation under practical conditions, the wide range of possibilities makes it impossible to find a figure of merit to cover all conditions. In both tone and programme tests a realistic compromise is necessary and some of the relevant considerations will be discussed in this Section.

It has been noted that the unwanted frequency fluctuations are generally accompanied by amplitude changes. In any recording system where reproduced output is proportional to velocity this must necessarily be so. The magnitude of the effect may be negligible, but a much larger degree of amplitude modulation can arise in other ways. It is produced, for example, by listening enclosures and even by the ear itself, whenever frequency modulation is present. Hence it is of interest to determine the subjective thresholds of both amplitude modulation and frequency modulation independently.

Listening experiments were carried out with both headphones and loudspeakers. The choice of loudspeaker can affect the results, especially in the case of frequency modulation of a tone carrier, where local irregularities in the loudspeaker-response curve produce undesired amplitude modulation. The production of harmonics by non-linear behaviour also tends to make the threshold value lower. Effects of this kind are, fortunately, largely masked when a programme source is used. The loudspeaker unit used for the experimental work was the B.B.C. standard unit, Type LSU/10, and experiments were conducted in the listening-rooms of the B.B.C. laboratories at Nightingale Square, London, and Kingswood Warren, Surrey, to simulate typical listening conditions. The reverberation time/frequency characteristics of these two rooms are fairly level at lower frequencies, but they fall gradually at high frequencies. The Nightingale-Square room, 2850 ft^3 in volume, has an average reverberation time of about 0·55 sec, and the Kingswood-Warren room, 1 890 ft^3 in volume, a reverberation time of about 0·45 sec. Experiments on a single subject suggested that this difference of reverberation time had no appreciable effect on the results.

(2.2) Tone Tests

In some initial experiments to determine the threshold of effects on pure tones, the signal was presented to the subject through high-grade moving-coil headphones. The headphones were subjectively calibrated for loudness level in the audio range, and the level chosen was 75 phons, which corresponds to the preferred peak listening level for light music programme material.[1] Various test tones between 50 c/s and 10 kc/s, modulated at frequencies unknown to the subject, were presented in random order. Various discrete levels of modulation of each test tone were presented, and the subject was asked whether the modulation was audible or not. Results with all subjects were found to be surprisingly inconsistent, and it was soon realized that aural fatigue and auditory imagery were serious factors in these conditions. The ear rapidly tired of the repetition of a pure tone, and once a given rate of modulation had been heard it gave rise to an after-image which was often most pronounced. Even if a pure tone was presented afterwards, a considerable "hysteresis" was often evident, the modulation still being heard in imagination.

Greater consistency resulted if the pure tone was presented first and the modulation gradually increased until the subject indicated that he was aware of a change. On account of the hysteresis effect, no attempt was made to repeat the test on a descending scale, and the tone was stopped immediately upon recognition of the modulation, in order to avoid fatigue. A new test tone then followed, together with a new modulation frequency. This procedure seemed to reduce fatigue considerably, and the aural imagery seemed also to be reduced by the subject's knowledge that the next rate would be different, even though he did not know how different.

The same method was also employed in other experiments, using a loudspeaker in place of the headphones, when it was not necessary to eliminate the effect of room eigen-tones.

(2.3) Programme Material Tests

In the experiments with musical programme material in listening-rooms, subjects were tested in small groups. It was impossible in these circumstances to arrange for the peak level to be exactly 75 phons for everyone, but the seating was arranged to give an equal scatter about this figure. In such rooms, however, the general loudness does not vary sharply with small changes of position and more serious variations probably arise from the non-axial frequency response of the loudspeaker.

Subjects who did not understand the terms "frequency modulation" and "amplitude modulation," and who were not well acquainted with their effects, were allowed to hear samples before the tests commenced. Each person had a switch which could be operated without the knowledge of others present on perceiving some disturbance thought to be due to the imposed fluctuation. This tended to reduce the competitive element and eliminate a chain-reaction after the most sensitive listener had recorded his impression. Individual identity was lost by this method, but identical groups could, of course, be reassembled and individuals tested singly when desired. The number of subjects who had operated their switches at any given time was registered on a meter indicating current flowing through the switch circuits.

The consideration of frequency fluctuations in a musical programme is complicated by the fact that there is no steady state of the carrier as in the case of pure tone. Recognition of the modulation of a pure tone depends on the ability of the ear to detect the addition of sidebands to a single line of the audio spectrum, whilst the other extreme—the recognition of modulation of white noise—is dependent on the ability to detect the order introduced into a chaos aurally invariant in time. A musical programme lies between these extremes, with the order/disorder relation continually varying. It is well known that a fixed degree of wow or flutter is more noticeable on piano music than on speech—indeed, a very large fluctuation is necessary sensibly to affect speech. Thus, to fix a permissible level of wow and flutter for practical purposes, it is best to select a type of programme material in which the effect is most obvious, so that on other material this level will be well within tolerance. The closest instantaneous approach to an ordered pure-tone regime is achieved by a piano or church organ. In many other instruments, as in the voice, it is common for the performer to introduce vibrato effects which may be mistaken for wow. In the main body of programme experiments, therefore, a piano solo was used, provided conveniently by an automatic player-piano. The instrument had to be well tuned, otherwise beats, greater than those inherent in the equal-temperament scale, could to some extent mask the wow.

One of the compositions used for experiments was Ireland's *Island Spell*, another was a Liszt transcription of Schubert's *Am Meer*, edited by removal of the more florid sections. Both works presented steady, slow chords, making possible the adoption of the following procedure. A particular waveform and frequency of fluctuation having been selected, the piano programme was modulated in increasing discrete steps of 10–20 sec duration and the reading of the switch meter noted at the end of each step. At any particular modulation level sufficient time was allowed for passages of easy wow recognition to be included, and variations of sound level and tone duration were reasonably constant between any group of modulation values. Any accidental weighting due to the character of the music when a certain

modulation level was presented could be corrected by giving further tests in which the process was repeated, with the same sequence in different musical context. The increase of wow was continued stepwise until the most insensitive listener indicated recognition. Suitable increments of fluctuation amplitude were decided by means of pilot experiments.

As the initial experiments progressed, it became apparent that the subjective threshold of the frequency fluctuations was sufficiently large to permit pre-recording of timed test sequences from the programme fluctuation generator. A check showed that there was no significant difference in the results given by the same group of subjects when tested direct with the generator and with a pre-recording done on a machine possessing inherent fluctuations of the order of 0·1% peak with waveforms of an irregular nature. Experiments with amplitude modulation were also carried out in this manner.

(3) GENERATION OF AMPLITUDE FLUCTUATIONS IN TONES AND PROGRAMME MATERIAL

To simulate the unwanted amplitude fluctuations in recorded programme material, it is necessary to amplitude-modulate a carrier (the programme) which has an extensive spectrum, and to do this in a manner which produces the necessary sidebands without significantly distorting the carrier or adding the modulation function to the output.

Fig. 1 shows the basic circuit employed, the programme- or

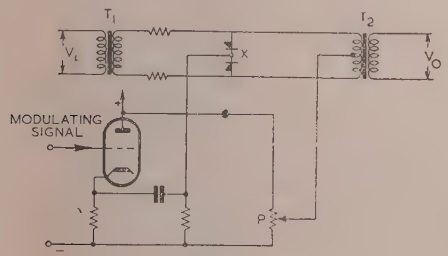


Fig. 1.—Circuit of amplitude-fluctuation generator.

tone-voltage being V_i and the modulated output V_0. The rectifiers X have resistances which are functions of the modulating e.m.f. applied to them. In addition to the resistive change, the modulating e.m.f. itself appears across X, but this can be eliminated by using the balanced arrangement shown, and the output from T_2 then contains only the wanted products. The modulating signal is fed from a cathode-follower to the junction of the rectifiers and mid-point of the primary winding of T_2, the potentiometer P serving to adjust the standing potential on the rectifiers to produce the best conditions. When the values of alternating and direct potentials are suitably proportioned, sensibly linear modulation up to about 25% can be obtained, but great care must be taken to avoid both resistive and reactive unbalances if the modulating signal itself is to be kept below the threshold of audibility. The calibration of this modulator presents no special difficulties.

(4) RESULTS OF AMPLITUDE-MODULATION LISTENING TESTS

(4.1) Tone with Sinusoidal Modulation

The subjects in this test were eight engineers fairly accustomed to listening to such effects, and the signals were presented through headphones at a constant level of 75 phons. The mean values of the results for threshold of detection are plotted in Fig. 2. The

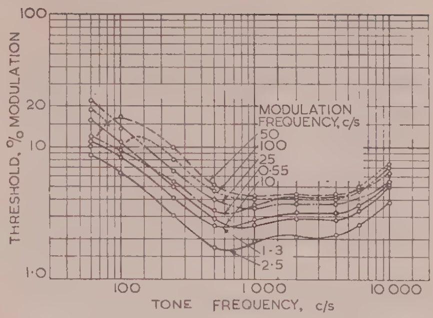


Fig. 2.—Subjective threshold for sinusoidal amplitude-modulation of tone.

general appearance of the variation is not unlike the Fletcher–Munson loudness contour[2] for moderate levels. There is enhanced perception of modulation frequencies around 3 or 4 c/s, but below 0·5 c/s perception becomes more difficult as memory is called into play. The finding that perception is most difficult at modulation frequencies around 50 c/s was at first suspect, but it confirms the results of Zwicker.[3]

The rise of threshold towards the lower test frequencies, which takes place at all modulation frequencies, is doubtless related to the decrease of just-noticeable loudness steps in the intensity gamut at low frequencies.[4] It will be observed that there is a sudden dip in the curve when a tone of, e.g., 100 c/s is modulated at the same rate. This is because the slow beating of frequencies in this condition is more easily detected than the higher-frequency components causing it.

Over a large part of the audio spectrum, amplitude modulation of 2–4% is necessary for aural detection, depending on the modulation frequency. The standard error of the means was of the order of 5% over this region, but it rose to about 10% in extreme regions, i.e. when using a modulation frequency of 100 c/s or a test tone of 10 kc/s modulated at almost any frequency. The means for a larger population may be even greater in the high tone-frequency region, for the sensitivity to high frequencies decreases with age and other factors, and in the present experiments the ages of subjects lay mostly between thirty and forty years. From these results it is evident that amplitude modulation produced directly in velocity-responsive systems is of relatively little consequence, for it will be shown that about 3% of frequency modulation is far more objectionable.

Some results were obtained, for a test tone of 1 kc/s only, showing the effect of amplitude modulation between 100 c/s and 10 kc/s. These revealed a steady reduction of threshold, except where detection is made difficult by beating effects, down to a minimum of 0·2% at 6 kc/s modulation frequency, after which the threshold rose to 0·4% at 10 kc/s.

(4.2) White Noise with Sinusoidal Modulation

The subjective effect of amplitude-modulated noise is of considerable practical interest. In conditions where the recorded signal/noise ratio is low, such as might occur in the dubbing of archives, it may well be a most significant factor. The results of experiments to determine the variation of subjective threshold for sinusoidal amplitude-modulation of white noise heard through a loudspeaker is shown in Fig. 3. As with pure tone, the most sensitive discrimination is found in the region of 3–4 c/s, where the threshold is 5%. At modulation frequencies higher than this, the threshold increases to 17·5%, where it remains up to the highest modulating frequencies employed. This difference from the pure-tone case may be explained by the difference in masking effects of noise and tone.

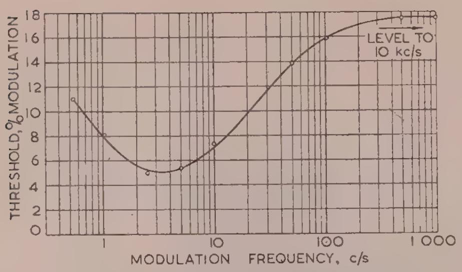


Fig. 3.—Subjective threshold for sinusoidal amplitude-modulation of white noise heard through loudspeaker.

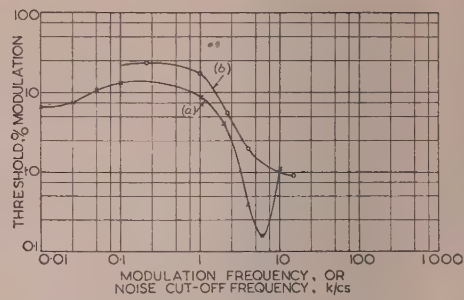


Fig. 5.—Subjective thresholds for h.f. amplitude-modulation of piano programme material.

(*a*) Sinusoidal modulation.
(*b*) Noise modulation.

(4.3) Piano Programme Material with Sinusoidal and Square-Wave Modulation

To determine the significant order of magnitude of amplitude modulation under practical conditions, the threshold of twelve subjects was determined when listening to piano programme material at the preferred level from a loudspeaker. As before, the subjects were engineers, to some extent familiar with the nature of the effect. The mean results for sinusoidal modulation, which are shown by curve (*a*) in Fig. 4, suggest that between 2 and 20 c/s the modulation frequency is not critical, whereas at 0·5 c/s the threshold almost doubles. At still lower frequencies the results appear to depend on the sense of absolute intensity.

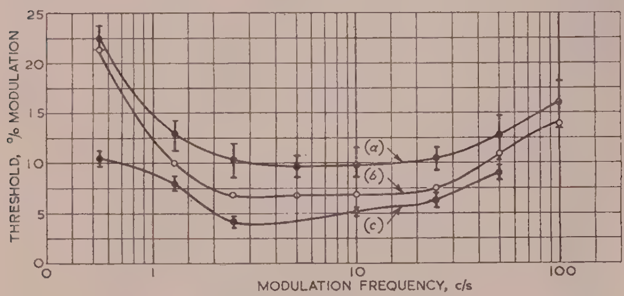


Fig. 4.—Subjective threshold for amplitude-modulation of piano programme material heard through loudspeaker.

(*a*) Sinusoidal modulation, population mean values.
(*b*) Sinusoidal modulation, single values.
(*c*) Square-wave modulation, population mean values.
I indicates standard deviation.

Curve (*b*) in Fig. 4, from observations taken under the same conditions, is for one particular subject, who was found throughout various tests to possess exceptionally acute discrimination.

Curve (*c*) in Fig. 4 shows the mean results of the whole group when the programme material was modulated by square-wave transitions. The shape is similar but the threshold, as would be expected, is depressed, particularly in the region of 0·5 c/s. The value there of 10·5% corresponds to about 1 dB change, popularly supposed to be the least noticeable change of level.

(4.4) Piano Programme Material with H.F. Sinusoidal and Noise Modulation

A great reduction of threshold is found when tone is amplitude-modulated at frequencies above the range 0–100 c/s detailed in Fig. 2. A test was carried out, on one of the more sensitive observers from the previous experiments, to ascertain if a similar effect existed for programme material. The results are summarized in curve (*a*) in Fig. 5. The general threshold is, not unexpectedly, considerably higher than for pure tone, but the variation is similar except for the peaks which occur in zones of beating in the pure-tone case. The rapid fall to 0·2% threshold at a modulation frequency of 6 kc/s corresponds to the result found at this frequency with 1 kc/s test material. In this region the aural impression is as though each percussive note excites a small tinkling bell, and it may not be aesthetically displeasing. It provides a good example of the hiatus between threshold and annoyance values of these parasitic effects. Such a modulation could be produced if a magnetic tape were finely and regularly milled across its width, but this is not a likely condition. However, similar random variations undoubtedly do occur and give rise to noise modulation. The threshold level was therefore examined when the programme was modulated with white noise restricted spectrally by a series of low-pass filters. The amplitude-modulation threshold is shown plotted against the noise cut-off frequency by curve (*b*) in Fig. 5. Except for the absence of the sharp dip at 6 kc/s, the results obtained closely resemble those of h.f. sinusoidal modulation. When modulated in this manner the programme is judged to have something of the character of low-grade tape-recording or of a slightly worn disc with a gritty accompaniment. The marked decrease of the threshold which occurs when noise components extend beyond 1 kc/s is interesting and indicates the value of polishing recording tapes.

(5) GENERATION OF PITCH FLUCTUATIONS IN PROGRAMME MATERIAL

(5.1) General

In Section 12.1 a mathematical analysis of wow- and flutter-generation is given in terms applicable to the generator now to be described, and to similar systems. The mechanical design of the generator is intentionally simple, in order that the inherent fluctuations due to imperfections of manufacture may be small. In addition, however, the analysis indicates that design parameters such as the relative disposition of the recording and reproducing heads and the speed of the recording medium, may be so chosen as to reduce even further the residual effect of manufacturing imperfections. It follows that subjective reaction to various controlled pitch fluctuations will not be influenced by the presence of appreciable accidental components. Calculations and measured values of the accidental components present in the output of the generator are given in Section 5.5.

(5.2) Description of Equipment

The mode of operation of the pitch-fluctuation generator is illustrated by Fig. 6. A rigid disc, composed of non-magnetic material, is supported in a plain bearing and rotated at constant speed by a synchronous-motor drive. The rim of the disc is coated with a mixture of magnetic oxides similar to that normally

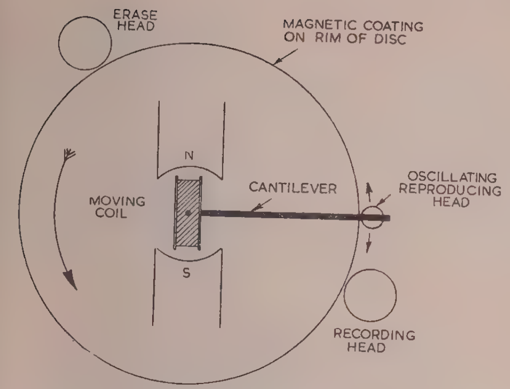


Fig. 6.—Sketch plan of pitch-fluctuation generator for programme-material experiments.

applied to the plastic backing of magnetic tapes. As the disc rotates it passes, in turn, an erase head, a recording head and a reproducing head. Each of these is separated from the magnetic coating by a gap of about 0·000 5 in. In the case of the recording head this separation from the coating results in a drop in recorded level which can be restored by a reasonable increase in the bias and recording currents. The output obtained when the gap between the reproducing head and the coating is d is proportional to the factor $\varepsilon^{-2\pi d/\lambda}$, where λ is the recorded wavelength. This represents a drop of approximately 55 dB in signal level for a separation equal to one wavelength. At short wavelengths the effect of even small separations is therefore very great, so that if the reproducing head is kept out of contact with the disc, the recording speed must be increased above normal to increase the wavelengths and reduce losses at higher frequencies. Furthermore, any variation in the nominal separation between reproducing head and coating, due to eccentricity or other causes, must be small in comparison with the shortest wavelength of interest, if noticeable amplitude-modulation of the signal is to be avoided. The choice of a high peripheral speed for the disc (in this case 100 in/s) thus permits both reasonable separation and sensible manufacturing tolerances. The reproducing head, which is described in detail in Section 5.4, is of light, miniature construction and is mounted on the end of a duralumin cantilever rigidly fixed to the centre of the moving coil in a conventional moving-coil system. The coil, and hence the cantilever and head, can move about the same vertical axis of rotation as the magnetic disc. Suppose that the disc is rotating and that a programme signal is fed into the recording head, together with the necessary bias current. If the reproducing head is stationary the recording track will be moving past it at constant speed, neglecting, for the present, inherent speed fluctuations due to imperfections in the manufacture of the wheel or its driving system. A l.f. alternating current fed into the moving coil will force it to oscillate about its axis, and the head will move to and fro. A wow or flutter is thus created in the reproduced output by the periodic fluctuation in the relative speed of the disc past the reproducing head. The frequency and amplitude of the wow or flutter depend on the exciting current and the length of the cantilever. Such a device is practicable only when the head is working out of contact with the magnetic medium. Since the moving coil is coaxial with the disc, the gap between the reproducing head and the coating remains nominally constant, whatever the amplitude of oscillation of the cantilever. Thus no amplitude modulation accompanies the flutter produced by the generator other than that due to any residual eccentricity of the disc itself. The calibration of the generator is described in Section 12.2.

(5.3) The Moving-Coil Drive System

The essentials of the electrical arrangement for the moving-coil drive are shown in Fig. 7. Response at frequencies as low as 0·5 c/s is necessary, and direct coupling is employed except for

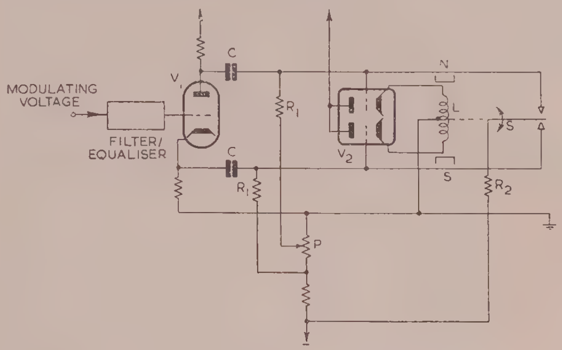


Fig. 7.—Circuit of moving-coil system of generator.

the penultimate stage. The centre-tapped coil L is fed in balance from the cathode-followers of the valve V_2. Preceding this stage is a phase-splitter V_1, into which the modulating e.m.f. is fed through corrective networks of suitable response. The balance of currents in each half of the coil can be controlled by the potentiometer P, which adjusts the bias on one output valve, and serves to centre the rotating head. If the horizontal cantilever attached to the coil were constrained to a centre position by mechanical compliances, points of resonance would tend to occur within the working range. Slow drift of the zero position is corrected by a simple step feedback system. A fine-spring wiper arm, S, is connected to the cantilever, and moves across two adjacent contacts separated by a thin sheet of insulation. The wiper is connected through a high resistance, R_2, to a negative bias source, and the two contacts to the grids of V_2. Should the coil drift from centre, the wiper touches one of the contacts so that R_2 is connected in parallel with R_1 across one of the grids. The potential on the grid changes with a time-constant dependent on the values of C, R_1, and R_2, reducing the current in one-half of coil L, and slowly restoring the central position. The correcting motion can be made sufficiently slow to be of no consequence to the output, since the degree of fluctuation is velocity-dependent and relatively hard to detect subjectively when the rate is low.

The disposition of the magnetic heads around the disc is of importance. It is shown in Section 12.1 that the recording and reproducing heads should be close together to reduce the effect of any accidental wow and flutter in the system. The reproducing head cannot be conventionally screened because mass would be added to the cantilever and the inertia would rise excessively. Extra precautions must therefore be taken in screening the recording and erase heads, and for this reason the erase head is moved to the opposite side of the disc, as far from the reproducing head as possible. The permanent magnet of the moving-coil system is also totally enclosed in a steel box to prevent leakage flux affecting the recording system.

(5.4) The Reproducing Head

For minimum inertia in the moving-head system the reproducing head must be as light as possible whilst maintaining adequate sensitivity for out-of-contact working. A conventional core consisting of a stack of Mumetal laminations would be too massive, and the core is therefore formed of a single rectangular lamination sprung into cylindrical form and held in position by a light clamp of plastic material. A reproducing gap is formed

between the two ends of the lamination, which are held a suitable distance apart by a spacer. Two coils are wound around the lamination and suitable spaces are cut in the plastic clamp to accommodate them. The coils are connected to the external circuit by fine wires which do not appreciably affect the mechanical stiffness of the system. The frequency response of this head can be made almost identical with that of good conventional heads, but its sensitivity is less.

(5.5) Accidental Pitch Fluctuations in the Generator

In Section 12.1 it is shown that the ratio of the instantaneous frequency, f_r, of the e.m.f. produced by the fluctuation generator to the frequency, f, of the input is given by the equation

$$f_r/f = 1 + (\omega v s/v_0^2)\cos\omega(t + T_0/2) - (v_1/v_0)\cos\omega_1(t + T_0)$$

where s is the mean distance between the recording head and the moving reproducing head,

v_0 is the constant component of the recording medium velocity,

$v \sin \omega t$ is a representative component of recording-medium velocity causing accidental flutter,

$v_1 \cos \omega_1 t$ is the velocity of the moving reproducing head, causing intentional flutter,

and T_0 is the time given by s/v_0.

If the generator is perfectly made then $v = 0$, and putting this value in the equation gives simply

$$f_r/f = 1 - (v_1/v_0)\cos\omega_1(t + T_0)$$

which is the desired result. In practice, however, v is finite, for there are always some small irregularities of traction even in well-designed systems. The irregularities are mostly of a random nature with superimposed periodic variations due to motors, idlers, etc.[5] The first equation above indicates how these accidental fluctuations can be minimized in the system described. The second right-hand term represents the uncontrolled variations, and it will be seen that the effect of a given value of v depends on the factor $\omega s/v_0^2$. For minimum accidental fluctuation in the output it is necessary to make the velocity, v_0, of the recording medium large, the separation, s, of the heads small, and the frequency, $\omega/2\pi$, of the accidental speed variation low.

In the fluctuation generator described, the peripheral velocity of the disc is 100 in/s and the radius is such as to give a fundamental value of ω of about 20 rad/s. A minimum value of $s = 1{\cdot}75$ in is set by head dimensions. The factor $\omega s/v_0$ gives the reduction of fundamental wow from any disc eccentricity as being $0{\cdot}35$. A greater reduction factor is unnecessary, for the eccentricity which can be tolerated is even more severely limited by the amplitude modulation which arises from variations in the separation between reproducing head and coating. More important are the speed fluctuations, at 25 c/s (the synchronous-motor speed) and other frequencies, from the driving system. These are eliminated by increasing the inertia of the disc by means of a flywheel and by driving it through a highly compliant belt. The resulting low-pass filter section leaves only fluctuations with frequencies predominantly in the region of 1 c/s, and these are effectively reduced by the correlation mechanism considered. With the same constants as before, except for $\omega = 2\pi$, the reduction factor is $0{\cdot}11$.

Actual measurements of accidental fluctuations from the generator gave a value of $0{\cdot}03\%$ peak, the fluctuation output consisting of a mixture of random components and periodic wow at the fundamental frequency of rotation in roughly equal parts. This small value is of no consequence from the subjective viewpoint.

(6) GENERATION OF PITCH FLUCTUATIONS IN TONES

A portable tone source[6] was used to produce wow and flutter in pure tones, with the modulating signal injected in place of the fixed warble tone normally produced in the instrument. In essence, the modulating signal was made to vary the gain of an amplifier controlling the capacitance presented to a tuned circuit, by means of the Miller effect, and hence to vary the frequency of the beat-frequency oscillator. Sinusoidal modulating signals at frequencies down to $0{\cdot}5$ c/s were generated externally by a Wien bridge oscillator, whilst square and pulse waveforms were produced by a multivibrator. The calibration of this system is described in Section 12.2.

(7) RESULTS OF PITCH-FLUCTUATION LISTENING TESTS

(7.1) Tone with Sinusoidal Modulation

The perception of frequency fluctuation is dependent on what is often called the differential pitch sensitivity of the ear, a property which depends on the conditions of experiment. The classic experiments of Shower and Biddulph[7] were made with a somewhat arbitrary waveform, in the sensitive region of 2 or 3 c/s, modulating pure test tones, which were then presented through headphones at various intensity levels. The experiments here described are intended to cover a wider range of modulating frequencies and waveforms at a fixed intensity level.

The mean results for a group of twelve subjects (engineers, aged 30 to 40 years) listening through headphones, are shown in Fig. 8. The graph indicates the peak fractional frequency-

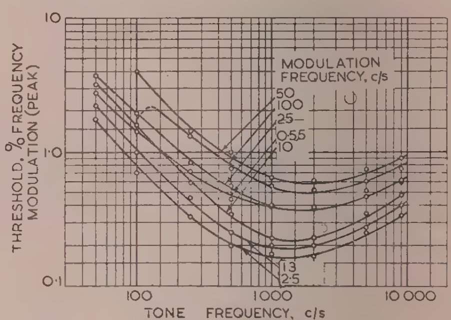


Fig. 8.—Subjective threshold for sinusoidal frequency-modulation of tone.

deviation just detectable at frequencies in the audio range, with the modulation frequency as parameter. When expressed in this manner the results show that much smaller percentages of frequency change than of amplitude change (Fig. 2) are noticeable. Other features are common, however, such as the increase of threshold at low frequencies, the sensitive region around 1 kc/s and the sensitivity to modulation frequencies of 2 or 3 c/s. There is also the zone of beating when 100 c/s is auto-modulated, and a threshold maximum around 50 c/s modulation frequency. The standard error of the means was of the order of 5% except at extreme values of both parameters, where it was twice this value.

Here again the experiment was extended, using 1 kc/s test tone, to include modulation frequencies up to 10 kc/s. Above the 0–100 c/s region, detailed in Fig. 8, the results show a remarkable similarity to those obtained for amplitude modulation (Section 4.1) if, instead of fractional frequency deviation, the modulation index is plotted, i.e. the ratio of deviation to modulation frequency. The threshold index shows a steady fall to a minimum value of about $0{\cdot}2\%$ at a modulation frequency of 6 kc/s and a subsequent rise to $0{\cdot}4\%$ at 10 kc/s. As before, this steady

variation is interrupted by zones where detection is difficult owing to beating effects.

(7.2) Tone with Square-Wave and Impulsive Modulation

The experiments described in Section 7.1 were repeated with square-wave modulation, and also with impulsive modulation produced by "differentiating" square waves in various degrees to produce exponential decays corresponding to time-constants of 40, 5 and 1 millisec. The greater number of sidebands produced by these more abrupt modulations of the tone would be expected to lower the threshold values. The results for square-wave modulation, shown in Fig. 9, confirm this and indicate that the threshold magnitude is also less dependent on modulation frequency than in the case of sinusoidal modulation. Modulation by impulses decaying with a 40 millisec time-constant gives threshold curves lying above the square-wave curves (though still below the corresponding sinusoidal thresholds) and dependence on modulation frequency is further reduced. For impulses with decays corresponding to time-constants of 1 to 5 millisec, the subjective threshold level is virtually independent of the time-constant or repetition rate, between 1 c/s and 25 c/s at least. The variation with these two parameters is of the order of experimental error, and a single representative curve has been drawn through the points. For square waves the standard error of the means was, in general, less than 5% in the 1 kc/s region, again rising to about twice the value in the 100 c/s region.

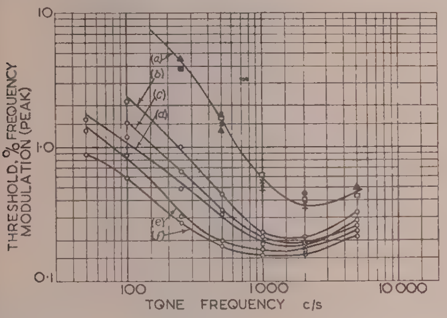


Fig. 9.—Subjective threshold for square-wave and impulsive modulation of tone.

(*a*) 1 millisec impulse: ○ 1·3 pulses/sec.; ● 5 pulses/sec.; × 25 pulses/sec.
5 millisec impulse: □ 1·3 pulses/sec.; + 5 pulses/sec.; △ 25 pulses/sec.
(*b*) 40 millisec impulse, 25 pulses/sec.
(*c*) 40 millisec impulse, 5 pulses/sec.
(*d*) 25 c/s square wave.
(*e*) 5 c/s square wave.
(*f*) 1·3 c/s square wave.

That the perception of a series of "plops" does not markedly depend upon the rate of presentation seems reasonable from spectral considerations, for at 5 millisec the envelope of sideband energy is spread beyond 60 c/s, and the detail of line spectrum within it is unlikely to be of first-order consequence to the filter system of the ear.

(7.3) Piano Programme Material with Sinusoidal Modulation

Possibly the most important fluctuation phenomena in recording systems are the periodic effects associated with non-centred discs or eccentric drive mechanisms. Modern recording practice involves a fundamental wow and flutter range from 0·55 c/s, found in long-playing and transcription discs, to 96 c/s, which occurs in ciné-film equipment. The experiments described in this Section were arranged to cover that frequency range as far as possible using the piano programme material previously discussed. The subjects, about seventy in number, were of various age groups and included fourteen females, sixteen males not

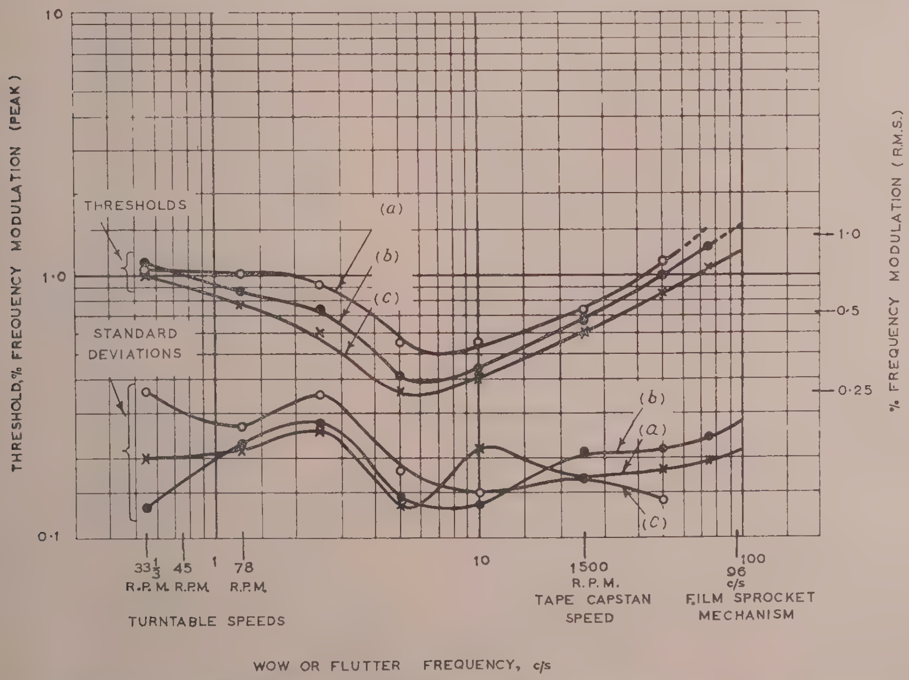


Fig. 10.—Subjective thresholds and standard deviations for sinusoidal frequency-modulation of piano programme material.

(*a*) 16 lay male subjects.
(*b*) 70 miscellaneous subjects.
(*c*) 14 female subjects.

associated with sound engineering, and over a dozen engineers having specialist interest in the experiments.

The results of the listening tests are shown in Fig. 10. The upper curves are the mean values; curve (*b*) for the whole population, curve (*a*) for the sixteen lay males, and curve (*c*) for the fourteen females. The lower curves, labelled correspondingly, show the standard deviations in each category. The curve of threshold values against number of subjects exhibiting a given threshold, i.e. the voting distribution, was reasonably normal for the larger population at all modulation frequencies. The histogram for 1·3 c/s modulation, typical of the results obtained,

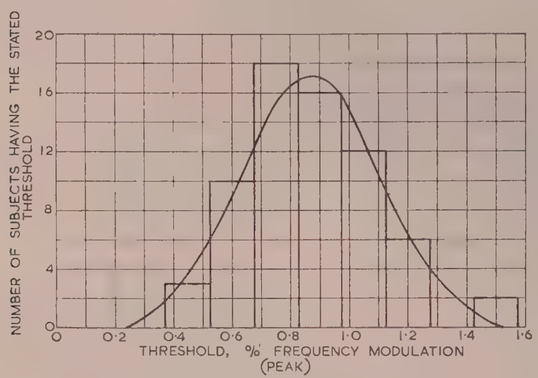


Fig. 11.—Distribution of subjective thresholds for 1·3 c/s sinusoidal frequency-modulation of piano programme material.

is shown in Fig. 11 with the corresponding normal distribution superimposed.

The standard deviations of the three populations, shown in Fig. 10, were calculated on a basis of normal distribution, although the threshold distributions of the two smaller populations were sometimes slightly skewed towards the higher fluctuation values.

The curves exhibit features in common with those for pure tone, except that the most easily recognized modulation frequency has increased from about 3 c/s to between 5 and 10 c/s, and the maximum which appeared round 50 c/s in pure tone tests is not now in evidence. A new feature is the slightly raised threshold at 2·5 c/s, which becomes progressively more marked with the increase of threshold values in passing from curve (*c*) to curve (*a*). A similar tendency is manifest in the standard deviations, which show a peak at 2·5 c/s in all three cases. The reason for this effect is not understood; it has been observed at various times, with different piano programmes and in different listening rooms. When the results of individual listeners are examined, two distinct classes appear, namely those with raised thresholds at 2·5 c/s and those without this feature. The absence of raised thresholds seems more common in subjects familiar with piano music, and the possibility of some rhythmical content confusing discrimination at this frequency, especially in lay subjects, cannot be dismissed. An examination of the fluctuations in sound intensity throughout the selected piano programme material showed that they were predominantly around 1 c/s, the fundamental rate of the music, falling sharply at frequencies above this, but there was relatively greater fluctuation around 2·5 c/s than the general trend would suggest. Although this aspect has not been pursued, it seems reasonable to suppose that such abnormal amplitude fluctuations might mask accompanying frequency fluctuations.

The results in Fig. 10 exhibit a very similar trend in the means of the populations represented by curves (*a*) and (*c*), despite the fact that with the relatively large deviations (of about 30%) there is little significance in comparing means at a particular modulation frequency. The fact that the curves closely maintain a similar shape over the whole frequency range, however, suggests greater significance. This was checked by combining the readings of the samples for all frequencies in the tests. To achieve this, the means and deviations at a particular frequency were first expressed as percentages of the whole population mean. Three new sets, including readings for seven frequencies up to 50 c/s, were then derived and their means and standard errors calculated for the combined number of measurements (i.e. for seven times the number of subjects in the original samples). The results are given in Table 1.

Table 1

Population	Mean	Standard deviation	Standard error
	%	%	%
(*a*) 16 lay male subjects ..	116	39	3·9
(*b*) 70 misc. subjects ..	100	32	1·5
(*c*) 14 female subjects ..	87	33	3·4

The threshold difference between subjects of samples (*a*) and (*c*), taking into account all modulation frequencies, clearly has a marked significance, although with such small samples there is a possibility of unsuspected bias.

(7.4) Piano Programme Material with Square-Wave Modulation

The tests involving square-wave modulation were restricted to fundamental frequencies below 10 c/s because of the limited response of the fluctuation generator. The group of subjects, 28 in number, was similar in composition to the whole population [curve (*b*)] of Section 7.3, and again the voting distributions were of normal character.

The results, summarized by Fig. 12, show that both the means

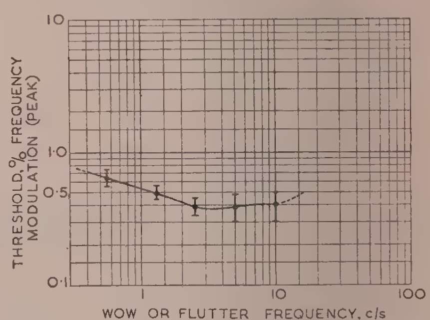


Fig. 12.—Subjective threshold for square-wave modulation of piano programme material.

● Population mean. I Standard deviation.

and deviations are, in general, less than for the sinusoidal case, as would be expected. The depression of threshold is most marked at the lowest frequencies, where the harmonics of modulation fall within a descending part of the corresponding sinusoidal modulation curve (Fig. 10), tending to enhance recognition. The reverse is true above 5 c/s, where the harmonic modulation frequencies are less effective in relation to the fundamental. With the spectrum restricted to about 100 cycles, the curves for square-wave and sinusoidal modulation would be expected to become nearly identical around 10 or 20 c/s, and this appears to be the case.

(7.5) Piano Programme Material with Impulsive Modulation

The effect of transient wow due to intermittent causes, such as tape adhesion, can be roughly simulated by applying pulses to the flutter generator. The recognition of a single pulse is not a matter substantially different from the recognition of a train of pulses in which the repetition rate is low, and for convenience the pulses were therefore presented at 2 sec intervals. The pulses reached peak amplitude after 20 millisec and decayed exponentially with a time-constant of 80 millisec. The same group of 28 subjects yielded a reasonably normal voting histogram, with a mean of 0·81% peak wow and a standard deviation of 0·21%.

(7.6) Piano Programme Material with Random Frequency Modulation

The residual wow or flutter of high-grade recording machines may not have obvious periodicity but may be predominantly stochastic in nature.[5] It is therefore of interest to examine the subjective effects of such fluctuations. The frequency fluctuations produced by the fluctuation generator when fed by a source of random noise yielded a Poisson form of distribution of the type which occurs in practical machines and which is described in more detail in Section 12.2.

In this particular experiment the piano programme material was modulated in discrete steps, and when the number of subjects showing a particular threshold was plotted against random modulation level, a histogram was obtained closely approximating to the type to which the random fluctuation amplitudes themselves belonged. It was found that the mode of the distribution was 0·2% (mean fluctuation) and the mean of the distribution 0·3%, which implies peak fluctuation of the order of 0·8%.

(7.7) Pitch Fluctuations in Other Programme Material

Vibrato of either pitch or intensity in a musical instrument tends to mask fluctuations caused by the recording system. From this viewpoint the piano, with its complete absence of vibrato, is possibly the most pure tonal generator available; its frequent use in the investigations described was based on this consideration. The organ may speak vibrato from the tremulant stop. An example of pitch-fluctuation threshold variation for light theatre-organ music is shown in Fig. 13, which gives the mean

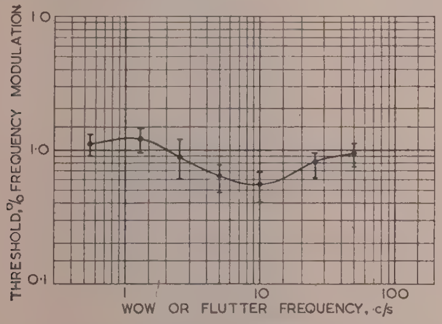


Fig. 13.—Subjective threshold for sinusoidal frequency-modulation of theatre organ programme material.

● Population mean. I Standard deviation.

results for nine subjects. As far as comparison with curve (*b*) in Fig. 10 is possible, the threshold is seen to be greater than that for piano programme material except at extreme flutter frequencies.

The singing voice. too, may have large degrees of vibrato, and, in general, the transient nature of speech sounds makes recognition of frequency fluctuation difficult. The threshold value for speech found by experiment was 2% to 3% peak at 1 c/s.

The perception of flutter in orchestral music is a very variable factor, for most wind instruments can produce vibrato and violins are particularly rich in the effect, with intensity vibrato of possibly 50% and pitch vibrato of about a quarter-tone at a modulation frequency of about 6 c/s. It is therefore hardly surprising that orchestral-music thresholds were found to be high. No detailed quantitative observations were made, since only the most critical case, the piano, is likely to influence recording-system design.

(8) EXAMINATION OF RESULTS

(8.1) General

The experiments described have provided information on the human discrimination of amplitude and frequency fluctuations of various waveforms and frequencies, on tone, noise and programme material. Considerable ranges of fluctuation magnitudes and frequencies have been explored in these experiments, and it is pertinent to inquire to what extent they may be encountered in practical systems.

(8.2) Amplitude Fluctuations in Practical Systems

With regard to fluctuations of amplitude, all frequencies in the audio band would seem possible. Consider, for example, a recording delay system such as is used in the programme flutter generator and is already in use for the production of artificial reverberation effects.[8] Any variation in the separation of the heads and the medium will cause amplitude modulation, and, according to the angular velocity of the drum, these variations may be at any frequency from below 1 c/s to about 25 c/s. Amplitude modulation at frequencies above this range to about 1 kc/s may be produced by creep or distortion in the drum material or by tool-chatter in the machining. Finally, small irregularities of surface finish or inhomogeneity of magnetic coatings or tapes may produce fluctuations up to the highest audible frequencies. These high-frequency amplitude fluctuations need be only of the order of 0·2% to be audible; in fact, some difficulty was at first experienced in eliminating such defects from the fluctuation generator.

(8.3) Pitch Fluctuations in Practical Systems

It has been noted (Section 7.3) that the fundamental frequency of pitch fluctuations in conventional recording systems may lie between 0·5 and 100 c/s. Fluctuations of about 0·1 c/s or less, known as drift, can be dismissed from the present consideration as subjectively unimportant unless the pitch wanders far enough and for long enough to be noticed by a listener with a sense of absolute pitch. Pitch fluctuation frequencies in excess of 200 c/s are unlikely to be generated to any extent in recording systems of current design with the values of inertia and compliance usually involved in the mechanical systems. There are, however, exceptions, for in magnetic recording systems the mechanical properties of the tape are such that it may be excited into longitudinal resonance by friction or random torques at frequencies of 1 kc/s or above, but these effects are unlikely to produce tape-speed changes greater than 1%.

(8.4) The Composite Curve of Pitch Fluctuation Thresholds

The subjective threshold for fluctuations at frequencies greater than 100 c/s was not measured on programme material owing to the limitations of the generator, but it is possible to assess its magnitude from other data. It has been observed in the case of pure tone that the threshold values of modulation indices for both amplitude and frequency fluctuations are virtually identical for modulation frequencies greater than about 100 c/s. From spectral considerations this is hardly surprising, for in this range

the identity of modulation indices expresses also the identity of energy distribution in the two principal sidebands, and the lack of phase-consciousness of the ear does not enable any difference to be resolved. If a similar identity is assumed for the aggregate of frequencies in the programme source, the frequency-modulation index for the threshold will follow the amplitude-modulation index experimentally determined for fluctuations above 100 c/s. The equivalent frequency excursion, Δf, may then be derived from the index value at any point by multiplying by $\omega/2\pi$, the modulation frequency. The new curve of relative Δf excursion above 100 c/s can then be normalized to produce the value actually observed at 100 c/s, and the composite threshold curve for the entire frequency range produced as shown in Fig. 14. A slight change of slope is noticeable at the 100 c/s junction, but no refined curve fitting was attempted because no more can be expected from this extrapolation than an order of magnitude.

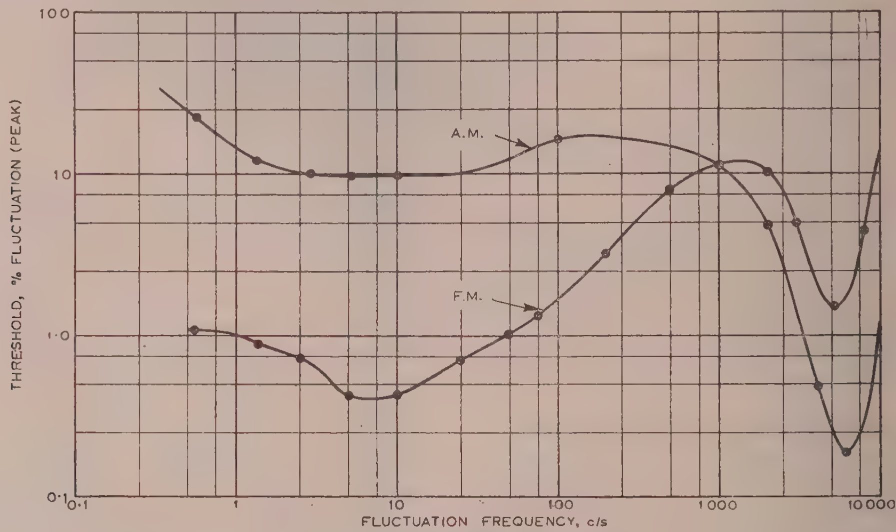


Fig. 14.—Composite curves of fluctuation thresholds (amplitude modulation and frequency modulation) for piano programme material.

The frequency-modulation threshold curve of Fig. 14 indicates that a minimum flutter of 1·5% is necessary to create a subjective effect above 100 c/s, and about 10% flutter is necessary in the region of 1 kc/s. There would seem to be little object in extending the range of flutter measurement beyond a few hundred cycles for programme recording systems as known at the present time.

(8.5) Specification of Tolerable Pitch Fluctuations

Turning to the problem of specifying tolerable fluctuations from the subjective viewpoint, it is clear that some compromise is necessary, for if, for example, 3 kc/s tone in a room were included within the practical scope, the permissible peak fluctuations could not exceed about 0·005%.[9] A reasonable compromise would be to consider the system fluctuations in the light of the more stringent programme conditions that occur most frequently in practice, such as the piano programmes used in the present experiments. A machine satisfactory on that basis would be adequate for organ and orchestral programmes, and virtually perfect for speech and any unusual transient effects. In other words, in the specification of tolerable fluctuations, and in their measurement, due weight should be given to the modulation frequencies to which the ear is most sensitive and the type of programme material on which fluctuations are most evident. A scheme for carrying out such subjectively weighted measurements will be discussed, but it must be realized that it may requir[e] supplementing under special conditions.

It is possible that when special conditions arise suitable deduc[-]tions can be made from results such as those given here, coverin[g] the subjective findings on tone and noise. As an example suppose a speech recording is made on a recorder having a 1·3 c/[s] wow, from a radio link which introduces a 9 kc/s whistle. Th[e] presence of the whistle may reduce the threshold wow level by a factor of about 50 : 1, for it is found that about 0·02% wow can be detected on the whistle alone when heard in an averag[e] room. The 9 kc/s level will be less than the 75-phon referenc[e] level used in the experiments, but as the amplitude-modulatio[n] threshold rises only by a factor between 2 and 3 for a loudness-level reduction from 80 to 30 phons, and the frequency-modulation threshold rises by a factor of about 2 for the same change, the wow threshold can be little more than twice the 0·02% value. Thus the tolerable wow level of 2% for speech alone is reduced, in this case, to approximately 0·04%. It may be argued that a modulation of an existing undesirable distortion is of no concern, but the example serves to illustrate the points at issue. In a similar manner amplitude modulation of background noise in a programme recording may indicate the presence of fluctuations which would not be detected on the programme alone. With such limitations in mind, the specification of tolerable fluctuations on programme material will be examined, special regard being paid to the formulation of a practical measuring scheme.

(8.6) A Universal Weighting Curve for Pitch-Fluctuation Measurements

The ideal subjectively weighted measuring device should, when threshold value is reached, give the same indication whatever the type and frequency of the fluctuation being measured. It will be shown that the results given here, as summarized in Fig. 14, enable a good approximation to this ideal device to be obtained. Consider first any single component of a sinusoidal wow or flutter lying in the range 0·5–200 c/s. If the fluctuation is applied to a discriminator giving faithful conversion in this range, and the output envelope is equalized by a network whose frequency response is the inverse of the frequency-modulation curve of Fig. 14, then the final output will be constant at threshold value throughout the frequency range. When there is more than one

fluctuation component present, the system will still suffice and the threshold will be indicated at the critical peak value of the weighted sum. If, for example, two equal components have frequencies close to one another, giving a slowly beating fluctuation, the threshold is essentially at the peak value of the mixture that equals the threshold of one component alone. If the frequencies are well separated, the one of lower threshold of course predominates. Thus, for a 1 kc/s tone, threshold is 0·2% at a wow frequency of 1·3 c/s and 0·4% at a flutter frequency of 25 c/s, whilst a check experiment showed that the threshold for an equal mixture is 0·26%. In the equalizing system suggested, half weight is given to the 25 c/s component and the resulting relative outputs are:

$$\text{At } 1\cdot3\,\text{c/s}\ 0\cdot2 \times 1 = 0\cdot2$$

$$\text{At } 25\,\text{c/s}\ 0\cdot4 \times \tfrac{1}{2} = 0\cdot2$$

$$\text{Mixture}\left\{\begin{array}{l} 1\cdot3\,\text{c/s}\ 0\cdot13 \times 1 \\ 25\,\text{c/s}\ 0\cdot13 \times \tfrac{1}{2} \end{array}\right\} = 0\cdot195$$

which, for practical purposes, is an identical and correct threshold indication in each case. The thresholds for several such mixtures were determined experimentally and found to agree equally well with the values which would be indicated by the measuring system.

Consider now the effect of another periodic waveform. Experimental observations have been obtained on square-wave modulation up to a frequency of 10 c/s. If square waves are applied to the suggested weighting network the resulting relative peak value will not alter substantially with frequency in the region of 6 c/s or above, but when the frequency is below this an overshoot takes place due to the selective increase of harmonic content. The amount of overshoot relative to the input is shown in Fig. 15 for modulation frequencies of 1 c/s, 2·5 c/s

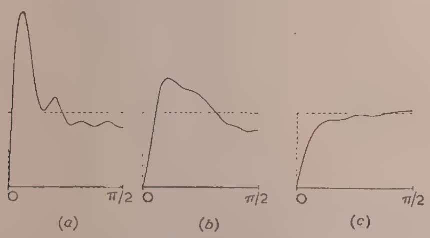


Fig. 15.—Response of suggested weighting network to square-wave input signal.

------ Input waveform. ——— Output waveform.
(a) 1 c/s square wave up to harmonic of 15 c/s peak output/peak input = 2·3
(b) 2·5 c/s square wave up to harmonic of 27·5 c/s peak output/peak input = 1·5
(c) 6 c/s square wave up to harmonic of 54 c/s peak output/peak input = 1·0

and 6 c/s. The graphs are approximate and show the effect of weighting the stated number of harmonics without considering relative phase differences. Constant peak-output of the network is achieved only if the square-wave input signal is reduced by the ratio of peak input to output, i.e. by approximately 1/2·3, 1/1·5 and 1 in the three cases considered. The subjective measuring device thus indicates the square-wave modulation threshold at fluctuation amplitudes reduced by these factors below those of the corresponding sinusoidal cases. The comparative square-wave thresholds as indicated by the proposed measuring chain and determined in the subjective experiments are given in Table 2.

The last two lines show that the agreement between the proposed indicating system and experiment is adequate, although

Table 2

COMPARATIVE THRESHOLDS FOR SQUARE-WAVE FREQUENCY FLUCTUATION

	Fluctuation frequency		
	1·0 c/s	2·5 c/s	6·0 c/s
Measured peak threshold for sinusoidal fluctuations, %	0·9	0·7	0·4
Calculated reduction factor for square-wave fluctuations ..	1/2·3	1/1·5	1/1
Peak square-wave amplitude when measuring chain indicates threshold, %	0·4	0·5	0·4
Measured peak threshold for square-wave fluctuations, % ..	0·5	0·4	0·4

there is over-compensation below about 1·5 c/s and under-compensation above.

Finally, consider the effect of the proposed weighting curve on the continuous spectra which are involved in the measurement of pulse and random modulations. The experiments yielded a mean threshold value of about 0·8% peak in each case, the bandwidth of the modulation being restricted to 100 c/s by the design of the fluctuation generator. The statistical distribution of amplitudes in the random-frequency fluctuation tests can be shown to obey a Poisson relation (see Section 12.2.1). It follows that amplitudes up to 0·8% peak comprise 99% of the fluctuations occurring in a random set of this nature. Assuming the maximum weighting to be unity at 6 c/s, where the sinusoidal threshold is 0·4% peak, the ratio of peak output to input should then be 0·5 in both the random- and pulse-modulation cases if true threshold is to be indicated. An approximate analysis suggests that such is the case. A modulating pulse of the form ε^{-kt}, $(t > 0)$ has a Fourier transform $(k + j\omega)^{-1}$, and its spectrum amplitude, $|k + j\omega|^{-1}$, is shown as curve (c) in Fig. 16 for an

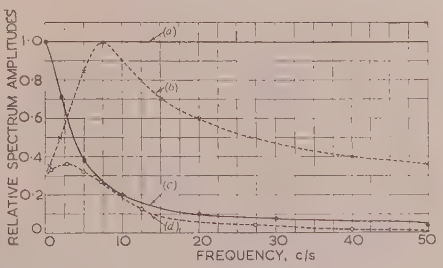


Fig. 16.—Determination of weighted output spectra for random and impulsive fluctuations.

(a) Reference line.
(b) Inverse threshold curve for sinusoidal modulation.
(c) Relative amplitude/frequency spectrum of impulse with 80 millisec time-constant.
(d) Relative amplitude/frequency spectrum of output signal.

80 millisec time-constant. The spectrum of the output signal [curve (d)] has been weighted by the inverse-threshold curve (b). The relative frequency spreads of curves (c) and (d) are similar, and by the Fourier-integral energy theorem the ratio of peak output to peak input should be roughly proportional to the ratio of areas beneath curves (d) and (c), which is about 0·6. In the random-modulation case, assuming an input spectrum level up to 100 c/s, weighting curve (b) is the relative amplitude spectrum of the output, and the peak-output/peak-input level will be the ratio of areas beneath curve (b) and line (a) in the range 0–100 c/s, which is about 0·4.

The suggested frequency weighting thus yields a constant

indication of subjective threshold value to within about 20% in all cases considered, which in view of the variations of judgment involved in such phenomena may be considered a reasonable result. An absolute calibration of the system is required, and this could be performed at a frequency of, say, 6 c/s, taking as standard the mean value of 0·4%. For critical standards it might be preferable to assume a value of the mean less twice the value of the standard deviation, so that fewer than 5% of listeners would be expected to notice a defect in a recording up to standard. This value corresponds to 0·14% peak, or 0·1% r.m.s., for a sinusoidal fluctuation, which is a value that has been accepted for many years as a good criterion for total permissible r.m.s. flutter in recording systems.

(8.7) A Universal Weighting Curve for Amplitude-Fluctuation Measurements

A similar method of weighting can be employed for amplitude fluctuations, using a curve which is the inverse of the amplitude-modulation threshold curve for piano programme material, and is also reproduced for the whole frequency range in Fig. 14. If information about h.f. modulation is not required, the peak around 6 kc/s can be removed; if it is retained, care must be taken to exclude sources of noise, such as thermal and shot noise, which are present even when the recording medium is at rest.

(8.8) Measuring System for Frequency and Amplitude Fluctuations

A block diagram of the proposed measuring system for both frequency and amplitude fluctuations is shown in Fig. 17. In practice, a tone from the reproducing chain in which frequency or amplitude modulation occurs would be fed, in turn, into the frequency- and amplitude-modulation chains and a reading obtained on the peak reading device or oscillograph, which had previously calibrated in absolute terms. A tone of frequency 3 kc/s, which is becoming standard for such tests, could of course be employed.

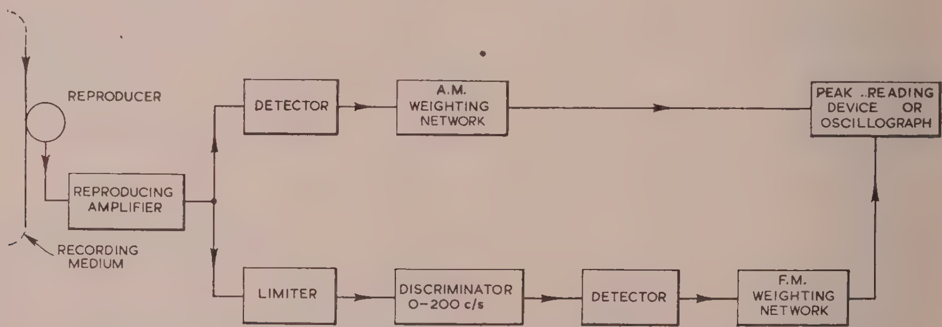


Fig. 17.—Block diagram of measuring system for frequency and amplitude fluctuations.

(9) CONCLUSION

The development of suitable test apparatus has enabled the subjective thresholds established in the course of the work described to be related essentially to the practical case in which a programme of music or speech is recorded for entertainment purposes. Experience has shown that there is a wide range of tolerances and subjective thresholds and that they depend on the nature of the programme being recorded. The recording of single tones, for example, may require exceptional manufacturing accuracy and adjustment in the recording and reproducing systems if no unwanted modulation is to be detected. The threshold in this case may also be confused by external factors concerned with the acoustics of rooms and loudspeakers. However, a practical and realistic approach to the problem is provided by considering results for the piano, for this instrument is well known to reveal, in the highest degree, the shortcomings of any normal programme-recording system. Consequently, a recording system manufactured to a standard suitable for piano programme material should be adequate for the recording of any other type of entertainment. The subjective tests carried out have involved a variety of listeners, expert and otherwise. Here again, such a compromise is necessary, for it would be possible to select individuals who are very sensitive to accidental amplitude- and frequency-modulations compared with the average listener. With these provisos the experimental results and calculations indicate that it is feasible to define frequency-weighting characteristics which can be incorporated into measuring chains so that the subjective effect of the accidental amplitude- and frequency-modulations occurring in any practical recording system can be registered.

It is hoped that the composite curves of measured fluctuation thresholds, summarized in Fig. 14, and the weighting characteristics which are the inverse of them, may provide a basis on which to lay down permissible standards of accidental amplitude- and frequency-modulation in recording systems. An encouraging feature of the results is that the standards required to suit the great majority of listeners are within reasonable engineering tolerances.

Many of the experimental results given here may have application in other fields such as those concerned with the mechanism of hearing.

(10) ACKNOWLEDGMENT

The authors wish to acknowledge the benefit of helpful discussions with many of their colleagues in the Research Department of the British Broadcasting Corporation. The paper is published by kind permission of the Chief Engineer of the Corporation.

(11) REFERENCES

(1) SOMERVILLE, T., and BROWNLESS, S. F.: "Listeners Sound Level Preferences," *B.B.C. Quarterly*, 1949, **3,** No. 4, p. 11; 1950, **5,** No. 1, p. 57.
(2) FLETCHER, H., and MUNSON, W. A.: "Loudness, Its Definition, Measurement and Calculation," *Journal of the Acoustical Society of America*, 1933, **5,** p. 82.
(3) ZWICKER, E.: "Die Grenzen der Horbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones," *Akustiche Beihefte*, 1952, **3,** pp. 125–133.
(4) RIESZ, R. R.: "Differential Intensity Sensitivity of the Ear for Pure Tones," *Physical Review*, 1928, **31,** pp. 867.

(5) AXON, P. E., and DAVIES, H.: "A Study of Frequency Fluctuations in Sound Recording and Reproducing Systems," *Proceedings I.E.E.*, Paper No. 799 R, January, 1949 (**96**, Part III, p. 65).

(6) MAYO, C. G., and BEADLE, D. G.: "Equipment for Acoustical Measurements," *Electronic Engineering*, 1952, **23**, p. 368.

(7) SHOWER, E. G., and BIDDULPH, R.: "Differential Pitch Sensitivity of the Ear," *Journal of the Acoustical Society of America*, 1931, **3**, p. 275.

(8) AXON, P. E., GILFORD, C. L. S., and SHORTER, D. E. L.: "Artificial Reverberation" (see page 624).

(9) ALBERSHEIM, W. J., and MACKENZIE, D.: "Analysis of Sound-Film Drives," *Journal of the Society of Motion Picture Engineers*, 1941, **37**, p. 452.

(12) APPENDICES

(12.1) Mathematical Analysis of Wow and Flutter Generation

Consider the recording system shown schematically in Fig. 18. The medium moves past a stationary recording head with a velocity compounded of constant velocity v_0 and a small

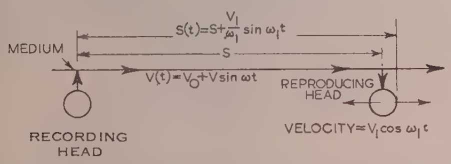


Fig. 18.—Mathematical relations in a frequency-fluctuation generator.

unwanted variation $v \sin \omega t$, a representative component of inevitable traction irregularities. Suppose that there is a correlation between the recording and reproducing process such that the reproducing head is centred at a distance s from the recording head, with reproduction taking place at a time of order $T_0 = s/v_0$ later. Let there be some relative motion of the reproducing head (as in the pitch-fluctuation generator described) which can be represented by a small velocity $v_1 \cos \omega_1 t$, thereby causing the separation of heads to be $s(t) = s + (v_1/\omega_1) \sin \omega_1 t$. It is proposed to determine the instantaneous frequency f_r which will be reproduced from the system under various conditions when a signal of frequency f is fed into it.

It is first necessary to determine T, the time at which reproduction takes place. Since both the velocity and distance traversed by a recorded element before reproduction are time-dependent, T is found from the equation

$$s(t+T) = \int_t^{t+T} (v_0 + v \sin \omega t)dt$$

Therefore

$$s + (v_1/\omega_1) \sin \omega_1(t+T) = v_0 T - (v/\omega)\,[\cos \omega(t+T) - \cos \omega t]$$

or $$T - T_0 = T_1 \sin \omega_1(t+T) + T_2[\cos \omega(t+T) - \cos \omega t] \qquad (1)$$

where $$T_0 = s/v_0,\ T_1 = v_1/v_0\omega_1,\ T_2 = v/v_0\omega$$

Clearly $$|T - T_0| \leqslant T_1 + 2T_2$$

Unless ω and ω_1 are very small, T_1 and T_2 are small quantities, since v/v_0 is of the order of 1% or less in any reasonable practical system and v_1/v_0 is of the same order in the flutter generator described in Section 3. Hence squares of these quantities may be neglected and T_0 substituted for T in the right-hand side of eqn. (1), yielding the approximation

$$T = T_0 + T_1 \sin \omega_1(t+T_0) + T_2[\cos \omega(t+T_0) - \cos \omega t] \qquad (2)$$

A recorded signal $V_i = \Phi(t)$ is reproduced as $V_0 = k\Phi(t-T)$, where k is some transfer constant, so that

$$V_0 = k\Phi\{t - T_0 - T_1 \sin \omega_1(t+T_0) - T_2[\cos \omega(t+T_0) - \cos \omega t]\}$$

The ratio of the instantaneous frequency f_r of output to the instantaneous frequency f of input will be $d(t-T)/dt$, by the usual definition, so that in this case

$$f_r/f = 1 - \omega_1 T_1 \cos \omega_1(t+T_0) + 2\omega T_2 \sin(\omega T_0/2) \cos \omega(t+T_0/2) \qquad (3)$$

If s is made sufficiently small,

$$\sin(\omega T_0/2) \simeq \omega T_0/2$$

and

$$f_r/f = 1 + (\omega v s/v_0^2) \cos \omega(t+T_0/2) - (v_1/v_0) \cos \omega_1(t+T_0) \qquad (4)$$

This is the equation which applies essentially to the high-speed disc and moving reproducing head used in the fluctuation-generator described.

In normal recording systems, however, the reproducing head is stationary and the velocity of the medium varies. Here, then, $v_1 = 0$ and eqn. (3) becomes

$$f_r/f = 1 + (2v/v_0) \sin(\omega s/2v_0) \cos \omega(t+T_0/2) \qquad (5)$$

Thus there is no frequency change when $\omega s/v_0 = 0, 2\pi, 4\pi$, etc., and there is a maximum change when $\omega s/v_0 = \pi, 3\pi, 5\pi$, etc. The head spacing may then be arranged, in the limit, either to cancel or double a particular flutter frequency, although if the wow is complex it is clearly best to make the spacing as nearly zero as possible to achieve best overall cancellation. In the usual description of a frequency-modulated wave an equation such as (5) can be regarded as expressing the instantaneous frequency. The form in the case of frequency flutter is

$$f_r = f + \gamma f \cos \omega t \qquad (6)$$

where the constant γ is independent of f.

The general expression for frequency modulation is of the form

$$V = V_0 \sin 2\pi[ft + (\Delta f/\omega) \sin \omega t]$$

where f is the carrier frequency, $\omega/2\pi$ is the modulation frequency, and Δf is the frequency deviation. This yields, for the instantaneous frequency,

$$f_i = f + \Delta f \cos \omega t \qquad (7)$$

Comparing eqns. (6) and (7) it can be seen that for flutter

$$\Delta f = \gamma f$$

i.e. the deviation is proportional to the recorded frequency. It is this identity which prevents the simulation of wow or flutter in complex programme signals by means of heterodyne or modulation systems. All such systems yield a deviation independent of the frequency of the signal components. The desired result can be achieved, however, by a memory device, which in the work described takes the form of a record on a magnetic drum.

(12.2) Calibration of Pitch-Fluctuation Generators

(12.2.1) The Magnetic-Disc Fluctuation Generator.

The degree of amplitude-modulation of a signal can be measured without much difficulty; the assessment of frequency deviation, however, requires care, especially when, as in recording systems, a component tone of the signal may be modulated at a rate in excess of both the deviation and frequency of the tone itself. This is a condition far removed from normal frequency modulation practice. Fortunately, wow and flutter are, by their parasitic nature, phenomena such that in practical cases the

frequency deviations are small. Thus the notion of instantaneous frequency can be usefully employed, and approximate information obtained on the nature of the resultant waveform.

The calibration of the magnetic-disc fluctuation generator was straightforward, since speed variations could be deduced from measurements centred on one particular frequency. For a given speed change the absolute frequency change in a recorded tone is proportional to the frequency of the test tone, so that a fairly high frequency was convenient for calibration. If too high a frequency was adopted, however, the general system noise and the amplitude modulation resulting from disc eccentricity created difficulties. A nominal calibration frequency of 3 kc/s was therefore chosen and any amplitude modulation present was removed by a simple diode limiter. The calibration was carried out by applying the 3 kc/s tone, in which wow or flutter had been introduced by the generator, to a discriminator network which transformed the frequency variations into amplitude variations. The disc fluctuation generator was not used to produce flutter frequencies in excess of 100 c/s, so the discriminator was not called upon to perform a linear amplitude conversion above this frequency. In this case a conventional phase discriminator of a type previously employed[5] was used to advantage. Measurement of the rectified output of the discriminator enabled the speed variations of the cantilever-head system to be related to the e.m.f. applied to the moving-coil drive system. The flutter region calibrated lay between 0·5 c/s and 100 c/s at magnitudes up to 1·5%, and the ratio of velocity-change to driving e.m.f. was found to be sensibly linear except for changes of less than 0·2% at frequencies below 1 c/s. In that region friction in the pivot system became an appreciable factor, but such conditions, being in practice below subjective threshold, were not investigated.

The calibration procedure for impulsive or square-wave modulation of the recorded signal was exactly the same, for the flutter waveforms were spectrally restricted to a few hundred cycles by the electro-mechanical behaviour of the system, and the frequency response of the conventional discriminator was still adequate. In certain experiments the cantilever-head system was driven by a source of random noise, and a pen recording was taken of the discriminator output. The frequency deviations were, in this case, of a stochastic nature, and the measurements were carried out (and results analysed) in the manner suggested by Axon and Davies.[5] The distribution of frequency deviations, as measured from the ordinates of the pen-recording, were closely described by a form of Poisson relationship,

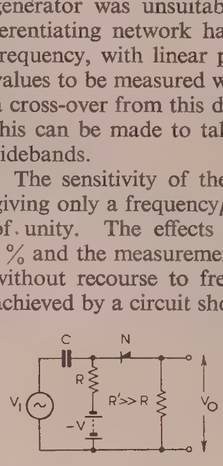
$$q(r) = \frac{n}{2A^3} r^2 \varepsilon^{-r/A}$$

in which A is a constant, $q(r)$ is the number of occurrences of a frequency fluctuation of magnitude r, and n is the total number of fluctuations recorded. The mean of this distribution has a value of $3A$ and the mode a value of $2A$. Thus the probable frequency excursion could be related to the level of noise applied to the drive system.

(12.2.2) **The Electronic Flutter Generator.**

Tests with pure tones involved a more extreme range of measurement, for test-tone frequencies ranged from 50 c/s to 10 kc/s, and modulation rates from 0·5 c/s to 10 kc/s. Here the type of discriminator used for calibrating the magnetic-disc generator was unsuitable. Use was therefore made of a differentiating network having an output simply proportional to frequency, with linear phase-shift, so enabling a wide range of values to be measured without error. All practical circuits have a cross-over from this desirable property at some frequency, but this can be made to take place outside the range of significant sidebands.

The sensitivity of the arrangement was unfortunately small, giving only a frequency/amplitude-modulation conversion factor of unity. The effects to be measured were of the order of 1% and the measurement of the signals had to be accomplished without recourse to frequency-dependent networks. This was achieved by a circuit shown in skeleton form in Fig. 19(*a*). The



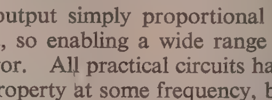


Fig. 19.—Discriminator circuit for higher range of flutter frequencies.
(*a*) Simplified diagram.
(*b*) Voltage relations.

signal V_1 after passing through the differentiating network CR was fed to the diode N, which cut off signals below a voltage V. This value was adjusted, as in Fig. 19(*b*), to be just below the minimum modulated input. The output voltage V_0 was thus modulated by a large percentage, and more easily amplified and measured. In view of the widely varying ratios of carrier to modulation frequencies that were possible, the measurement was conveniently carried out by means of a scale on a cathode-ray oscillograph, displaying V_0 on a suitable time axis. Without this visual display there were measuring difficulties when the modulation frequency approached, or exceeded, the carrier frequency.

621.385.032.216

Paper No. 1790 R
Apr. 1955

A STUDY OF THE LONG-TERM EMISSION BEHAVIOUR OF AN OXIDE-CATHODE VALVE

By G. H. METSON, M.C., Ph.D., M.Sc., B.Sc.(Eng.), Associate Member.

(*The paper was first received* 17*th June, and in revised form* 15*th November*, 1954. *It was published in April*, 1955, *and was read before the* RADIO SECTION 20*th April*, 1955.)

SUMMARY

The problem of maintaining satisfactory emission in an oxide-cathode valve over long periods of time is examined in broad rather than detailed manner. The approach is essentially experimental and starts from the basic assumption that the cathode will maintain its emission so long as it possesses a sufficiency of metallic barium in its oxide matrix. The general plan of the paper is to examine the several factors leading to loss and gain of barium metal within the cathode and then to attempt to estimate the conditions likely to result in a credit balance of barium over a long period of time. The tentative conclusion reached is that a cathode working under a current load in gas-free surroundings will continue to emit until the barium-oxide matrix is exhausted by electrolysis and evaporation.

Part 1. EMISSION BEHAVIOUR OF AN OXIDE CATHODE ON A PLATINUM CORE

(1) INTRODUCTION

(1.1) Scope of the Problem and Manner of Approach

The long-term emission stability of a high-vacuum oxide cathode is a subject which has received little attention in the literature. The rise of submarine telephony in recent years has, however, focused interest on the phenomenon, and the present paper attempts to assess the degree of probability of very long emission life in a cathode running under favourable conditions. Partly owing to the complexity of circumstance wherein the cathode works and partly owing to the time element involved, the problem of assessment is a formidable one, and the scope of the paper at this stage is perhaps over-ambitious. A certain modest progress has, however, been made over the past four years, and from this has emerged a hypothetical emission-life model which may be of interest to some engineers.

The approach to the problem has been twofold. The first line is the conventional one of making pentodes under the best conditions known and subjecting them to prolonged life test. The disadvantages of this technique are obvious—the tests are time consuming and lead to little appreciation of the complex interplay of forces contributing to the observed behaviour. Such tests have, however, a part to play, and some use will be made of them in the paper.

The second line of approach is a more difficult but potentially more fruitful one. An attempt is made to analyse the several forces in the valve leading to emission decay and emission enhancement (activation). A synthesis of the two opposing tendencies is then undertaken in order to find which set is in credit balance. The paper is primarily concerned with developing this analytical approach.

It has been found convenient to divide the paper into two parts. Part 1 examines in some detail a general pattern of behaviour of an oxide cathode supported on a chemically inert core metal (platinum). It has been found that the elimination of oxidizing and reducing chemical actions of the core metal is a help towards understanding the basic behaviour of the cathode.

Part 2 deals with the case of nickel-core cathodes. The simpler example of pure nickel is first examined, and this is followed by the complex active core of the type used in industry.

(1.2) Valve Life

In a recent paper by Eaglesfield[1] the conclusion is reached that the primary cause of performance deterioration in a well-processed repeater valve is the growth of resistive interface. Another paper[2] examines common receiving valves and concludes that deterioration is due partly to interface-resistance growth and partly to gas de-activation of the oxide cathode.

It is some four years since these papers were prepared, and during that period a degree of progress has been made in the processing of high-grade valves—the interface effect can be eliminated[3] and considerable steps have been taken to reduce residual gas effects. Such valves of the high-slope pentode type show a negligible decay of mutual conductance over the first ten thousand hours of life.

In ordinary circumstances it would now seem reasonable to allow the life problem to mature naturally—the next five or ten years should consolidate a situation which is already showing promise of coming under control. However, present circumstances are peculiar in one respect; the expanding submarine-telephone art is largely dependent on valve life, and we are looking to the valve maker for something more than a considered optimism.

(1.3) Cathode Life

Forethought in design and care in production can probably eliminate the ordinary forms of valve failure with the one exception of cathode emission decay; valve life therefore reduces to the problem of cathode emission life.

The conventional barium–strontium-oxide cathode requires for its essential functions of matrix conductivity and surface emission a stoichiometric excess of the metal barium. This barium activator is subject to various forms of wastage, of which one, i.e. evaporation, is inexorable. An emission life of indefinite length is therefore only compatible with the existence of a perpetual barium regenerative action.

(1.4) Life Model

The present work is based on the use of a dynamic life model which asserts a continuous loss of barium from the barium–strontium-oxide matrix and a continuous replenishment. It is assumed that there are three possible ways in which the excess barium is lost and two possible ways in which it is regenerated.

The course of the paper follows the general order of Table 1—the "loss" processes are first examined and then the "profit" processes, and subsequent Sections attempt to draw up a profit-and-loss account in barium for practical cathode cases. If the barium account shows a continuous decrement, the cathode life is finite; if the decay ceases in a finite time at an adequate level

Dr. Metson is in the Post Office Engineering Department, Research Branch.

Table 1

Loss processes	Profit processes
(*a*) Oxidation of barium by residual gas (*b*) Loss of barium to core metal (*c*) Evaporation of barium from emitting surface	(*d*) Electrolytic dissociation of basic matrix BaSrO (*e*) Chemical reduction of basic matrix BaSrO

of barium balance, life is limited only by some factor other than cathode activity.

(1.5) Core Metals

Three types of core metal have been used in the course of the work, with the object of providing either the presence or absence of a reducing agent in the cathode system. The active or reducing core metal is a nickel alloy containing 0·05–0·10% of magnesium and silicon—a typical alloy used by manufacturers for receiving valves. The two other core metals are pure nickel and pure platinum. The analysis of these so-called pure metals is undertaken spectrographically and is a matter of some conjecture. The platinum analysis, for example, gave the following figures:

Magnesium	0·0001%
Silicon	0·0005%
Copper	0·0005%

(1.6) Influence of Core Metal on Initial Activation

It has been shown recently[4] at the Post Office Research Station and in the United States by Rouse and Forman[5] that there is no significant difference in the initial emission level of cathodes processed on active nickel or pure platinum cores. Under the best laboratory conditions both core metals give rise to cathodes with emissions of 2–10 amp/cm^2 at 1 020°K. It seems, therefore, that the initial emission is independent of the amount or nature of the core activator. Emission decay rate will be shown later to be a function of initial emission level, and it is therefore a convenient circumstance that a standard diode fitted with different core metals can be initially activated to a common level.

(1.7) Experimental Plan

. The experimental plan has been indicated in Table 1, and it remains to see how core variation can be fitted into this scheme to the best advantage. In investigating the "profit" processes it is essential to separate the electrolytic and chemical-reduction actions, and this is attempted by using core variation. To examine, for example, the electrolytic process the inert platinum core is used; profit in barium is then examined under conditions of current load and no-load through the cathode. Chemical reducing action is examined by employing the active nickel core under conditions in which the possibility of electrolytic action is absent, i.e. with zero cathode current.

As the experimental work has involved the preparation and testing of a large number of standard diodes it has been found convenient to undertake the bulk of the work with one core type; to try to recognize some pattern of behaviour; and then to examine differences with the other types. At the time of starting the investigation only two types were available—active nickel and pure platinum—and the choice was therefore restricted to these alternatives. It was decided to undertake the general investigation on platinum for the following reasons. Use of platinum would avoid the complications following the inevitable growth of a resistive interface layer; the inert metal would enable the electrolytic effect to be investigated without the confusing presence of a reducing agent; and finally there was an instinctive feeling that the platinum case was the simplest and most fundamental form for a basic examination. As matters have turned out there have been no regrets on this decision.

Before turning to the emission-decay effects themselves, a short Section will be devoted to a description of the basic experimental tube and the technique of emission measurement.

(2) STANDARD DIODE AND METHOD OF EMISSION MEASUREMENT

(2.1) Standard Diode Type 6D15

The standard diode type 6D15 employed in the work has a 2-watt rectangular-box indirectly-heated cathode and a collector consisting of a fine molybdenum-wire grid structure on parallel copper or nickel support rods. The structure is, in fact, identical in general form to the cathode and control-grid structure of a typical high-slope pentode. The adoption of such an assembly has many advantages—it can readily be expanded to triode or pentode form, it enables the total emission of diodes and pentodes to be directly compared, etc.

The voltage/current characteristic of the diode is shown in Fig. 1, and it will be observed that a current of 100 mA can be

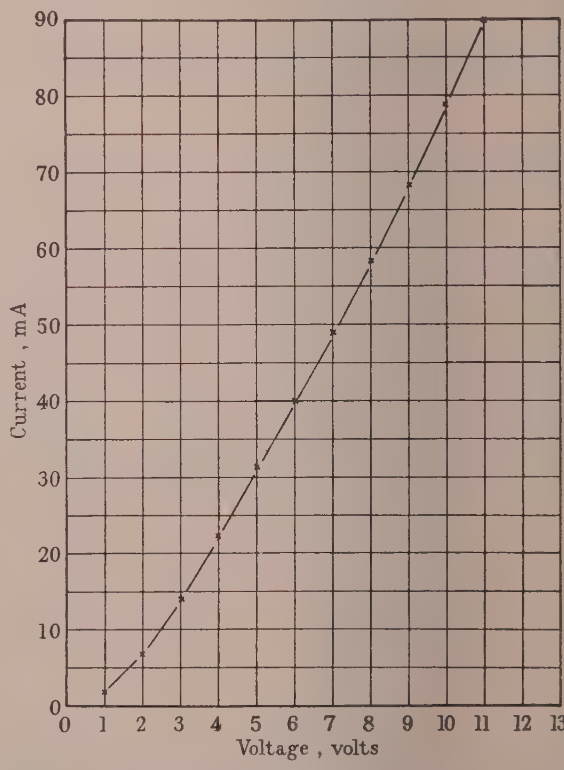


Fig. 1.—Typical voltage/current characteristic of type 6D15 diode.

drawn from the cathode surface (0·5 cm^2 area) before an ionizing voltage is encountered. The temperature of the cathode is related to the heater voltage in the manner of Fig. 2.

The structure is inherently simple in construction, has a minimum number of piece parts and the collector is open to efficient r.f. heater treatment.

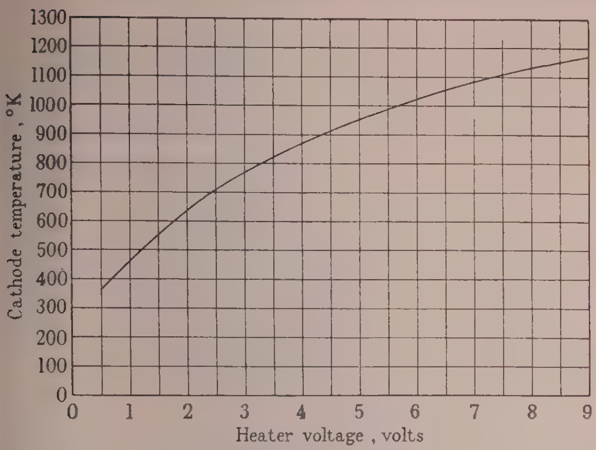


Fig. 2.—Cathode-temperature/heater-voltage characteristic of type 6D15 diode.

(2.2) Processing Procedure

The diodes were processed on high-grade bench pumps under the best laboratory conditions available with present-day equipment. Details of the pump schedule are given in Section 8.1, from which it can be seen that the valves are activated on the pump. Manifold pressure after breakdown was required to be better than 1×10^{-6} mm Hg. Gettering followed normal commercial practice, except that it was on a rather generous scale (6 mg of barium).

One point of interest may be seen from Section 8—all core types are given identical treatment, and this involves the passage of a heavy current through the cathode. This current is beneficial to the active-core type and essential to the passive type.

(2.3) Measurement of Total Emission

The total emission of a cathode is directly proportional to the matrix conductivity, which in turn is a direct function of barium concentration within the matrix. It follows that the measurement of total emission is a measurement of the barium content of the cathode matrix. The measurement technique employed in the present work is the low-temperature (700° K) method. With the type 6D15 diode a cathode temperature of around 700° K is obtained by injecting 400 mW into the heater circuit, and the total emission is measured by applying a voltage of +5·0 volts to the collector. The total emission should be 3–10 mA in a well-activated tube at the beginning of life. The reproducibility of the method depends on the known degree of resistance stability of the tungsten heater and the accuracy of measurement of the 400 mW of heater power. Both of these practical points were satisfactorily arranged.

All total emission measurements in the present work were made by the low-temperature method, but it is of interest to quote the proportionality factor for emission on the diode type 6D15 at 700° K and the normal operating temperature of 1 020° K. The factor varies from valve to valve, but the average figure of 1 000 is representative. This figure, which has been obtained by Holmes, converts total emission in milliamperes at 700° K to total emission in amperes per square centimetre at 1 020° K.

(2.4) Assembly of Test Batches

The measurement of a total-emission decay effect has always been undertaken by assembling a batch of samples, measuring the individual decay curves, and presenting the result as a mean characteristic. The method is laborious, but it prevents the collection of misleading information from individual "rogue" valves. The batch size varies somewhat according to the importance of the measurement and the availability of samples, but is normally between six and ten valves.

The assembly of a batch requires care, as it will be shown later that decay rate is a function of emission level. The aim, therefore, in batch assembly is the selection of samples with a closely grouped set of total-emission values and with a mean value as high as practicable. An initial mean total emission of 6–7 mA is typical. A histogram of a group of diode type 6D15 samples is shown in Fig. 3, and it is clear that the selection of batches

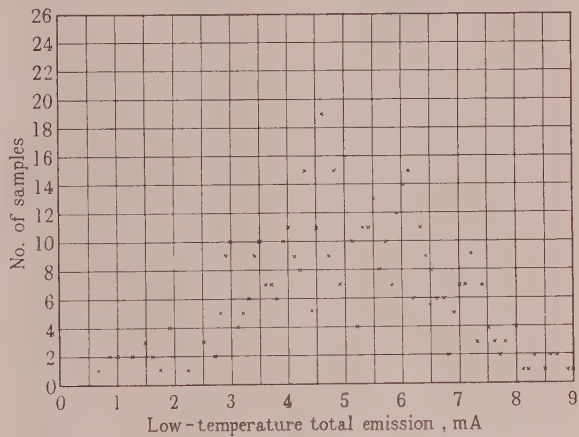


Fig. 3.—Distribution of total emission.

with mean total emission of 6–7 mA is somewhat wasteful, but it has the advantage of extending the range over which decay occurs.

(3) THE BASIC LONG-TERM DECAY EFFECTS ON A PLATINUM-CORE CATHODE UNDER ZERO-LOAD CONDITION

(3.1) The Zero-Load Decay Characteristic

In Fig. 4 is shown the zero-load decay characteristic of a typical group of platinum-core type 6D15 diodes running at a cathode temperature of 1 020° K. The mean total emission has been plotted on a logarithmic time scale to emphasize the fact that after 5 000 hours there is still no sign that the decay has come to an end. The rate of decay is, however, continually diminishing, and this is brought out in Table 2, which has been derived from Fig. 4.

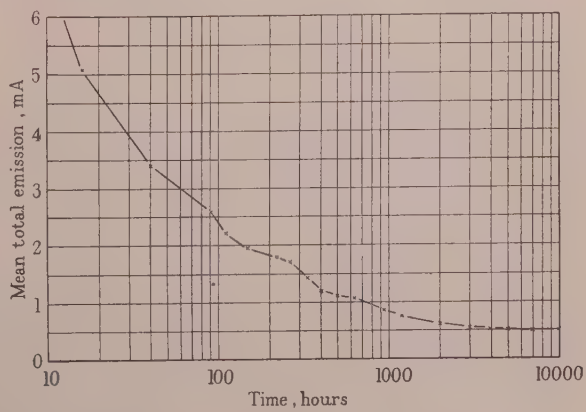


Fig. 4.—Zero-load decay characteristic.

Table 2

Time of life	Mean rate of decay
hours	μA/hour
5	200
15	70
25	50
75	16
150	6·0
450	1·4
750	0·60
1 500	0·22
2 500	0·08
4 500	0·02

(3.2) Influence of Initial Emission on Zero-Load Decay

The decay characteristics of a typical pair of valves selected for their difference in initial total emission is shown in Fig. 5. The

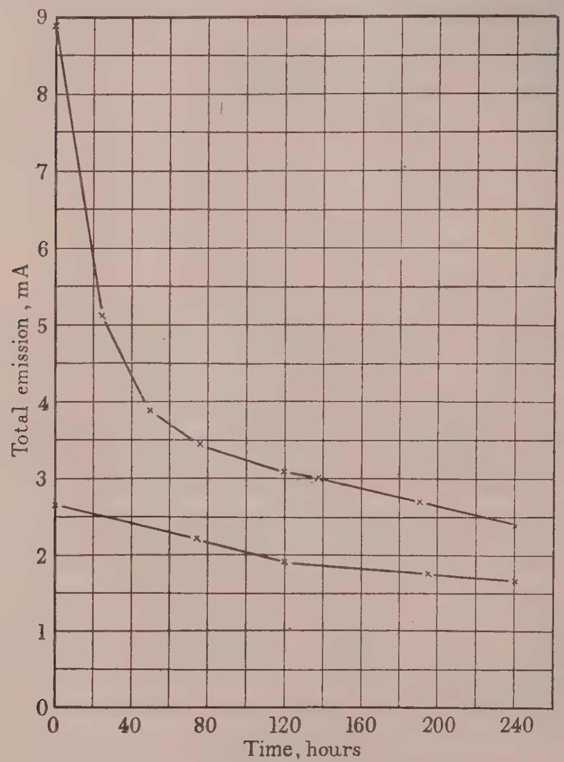


Fig. 5.—Influence of total emission level on decay rate.

curves show that the rate of decay at any time is related to the magnitude of the emission at that time. The results emphasize the necessity for caution in assembling groups of valves for determination of a mean decay characteristic (see Section 2.4).

(3.3) Influence of Cathode Temperature on Zero-load Decay

The influence of temperature on the zero-load decay rates of groups of platinum-core cathodes is shown in Fig. 6. No decay occurs in the temperature range 650–800°K in 250 hours, and this period has been extended to 1 000 hours with similar results. Above 800°K decay sets in and thereafter increases rapidly with increase of cathode temperature.

The results are set out in another form in Fig. 7, which relates

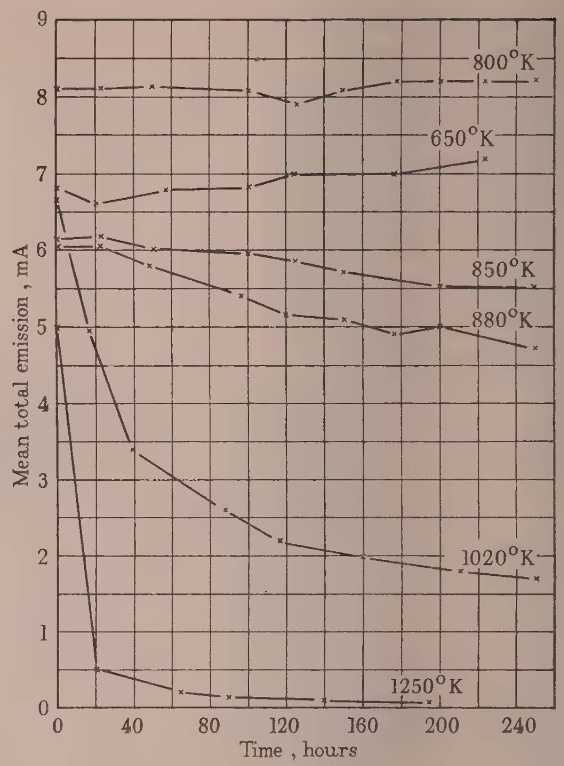


Fig. 6.—Influence of cathode temperature on zero-load decay rate.

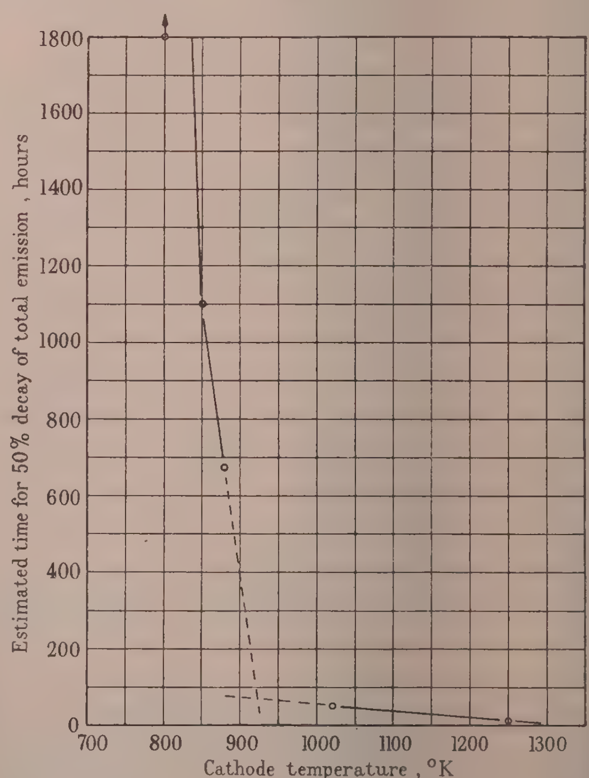


Fig. 7.—Relation of cathode temperature to time for 50% emission decay.

cathode temperature to the estimated time in hours to decrease the total emission by one-half of its initial value. It seems that there is a fairly narrow temperature range below which the emission is stable with time and above which a progressively rapid decay occurs.

(4) CAUSE OF DECAY ON A PLATINUM-CORE DIODE UNDER ZERO-LOAD CONDITION

(4.1) Survey of Possibilities

The decay effects described in Section 3 are assumed to be due to loss of excess barium by the cathode matrix. This loss must have been caused either by destruction of barium *in situ* by residual gas action, or by physical removal of barium in elemental form from the matrix. This removal may occur by direct evaporation or by core solution. These are the logical possibilities, and it is probable that all three modes of loss contribute to the decay phenomenon in a practical valve. It is the object of this Section to attempt to place the three modes in a relative order of magnitude, for the type 6D15 diode.

(4.2) Investigation into Residual Gas Destruction of Excess Barium

Evaporation and core solution of barium from a matrix are specific physical operations and may in time be defined in terms of mechanical dimensions, temperature, solubility, etc. The two quantities can be regarded as constant in similar cathodes running under similar conditions. Gas attack is, however, a somewhat random operation depending on the fortuitous circumstances of valve manufacture. The decay characteristic of Fig. 4 is therefore compounded of two fixed physical components and one of a random nature. In starting the present inquiry into the magnitude of gas action contributing to the decay of Fig. 4, it becomes essential to define the base-line of processing quality control on which the residual gas pressure depends. From such a base-line a series of controlled sorties can be made to influence the decay rate by further improvements in the degassing treatment of individual members of the valve structure.

The base-line of process control adopted for the present investigation is that used for the preparation of platinum-core pentodes for Post Office submerged repeaters (see Section 8.2). The process is based on hydrogen firing of metallic piece parts prior to assembly, and its success depends on a correct assessment of factors such as firing temperature, duration of treatment, hydrogen purity, etc. The results of this standardized processing have given satisfaction with pentode production, and the system is probably the best available starting-point for the diode type 6D15 investigation. The results of Section 3 were obtained on diode type 6D15 samples pre-processed in this manner.

The way to go is now clear. The batch decay rate of standard processed diodes is used as a control against which to compare the decay rate of an experimental batch having standard processing for all piece parts except one, which is given a much more elaborate degassing treatment. By comparative experiments of this type the whole of the metal components are systematically examined.

An exhaustive programme of work on the above lines failed to alter the decay rate over 250 hours by any greater amount than was to be expected from experimental error. These results seem to indicate either that residual gas is not diffusing from the metal components of the diode, or at least that such gas is generated too slowly to influence the decay rate in a measurable way over the period of observation. The batch size in these experiments was six to eight samples. A typical result is shown in Fig. 8.

Attention was next turned to the possibility of gas evolution

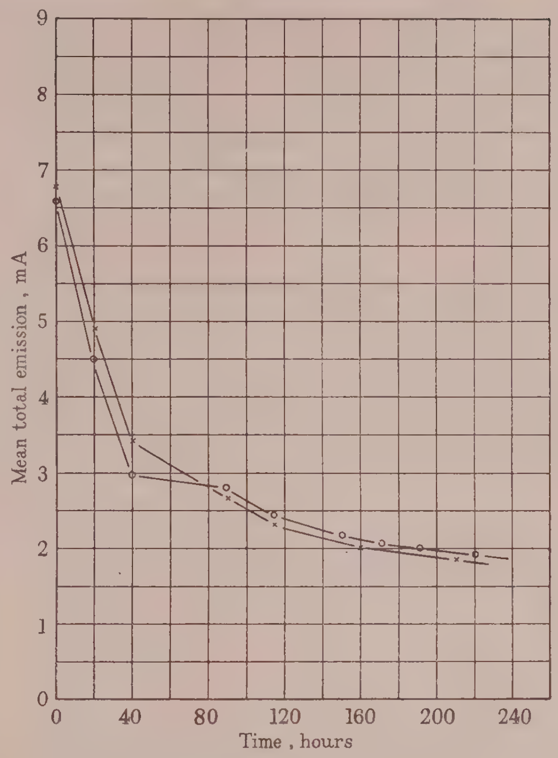


Fig. 8.—Decay characteristic of a control and experimental batch of type 6D15 diodes.
× Control batch.
○ Specially outgassed collectors.

from glass and mica, and certain of these experiments are perhaps of sufficient interest to merit brief description. Two examples of the glassware investigation will be mentioned. The normal running temperature of the envelope is 35°C, and one experimental batch was run in an oven at 100°C. No difference in decay rate over 250 hours was observed between the experimental batch and a control batch running at the normal envelope temperature. In a second experiment the area of the barium getter flash on the interior surface of the envelope was varied. The approximate surface coverage in the control case was 20%, and in the two experimental batches it was 40 and 80%. No difference in decay rate was observed. From such experiments it was concluded that the main decay effect was not due to the evolution of gas from the glassware of the diode type 6D15.

The mica problem was approached in two ways—first the thermal vacuum pre-processing was altered over a substantial range, and secondly the quantity of mica used was altered by one-half. Neither variation affected the decay rate over 250 hours.

One final experiment may be mentioned, since it indicates the extent of the work devoted to the gas investigation. It was felt that there was a possibility of electrolytic oxygen evolution from the hot alumina used as tungsten-heater insulant. To test this hypothesis a new heater was designed with thicker tungsten wire and so proportioned that, for the same heater power and cathode temperature, the new heater ran at approximately 100°C lower in tungsten temperature than the standard heater. As electrolytic conductivity is an exponential function of temperature, this difference should make a substantial change in the rate of oxygen evolution if this does indeed occur. Comparative tests of decay rates, however, gave a null result, and alumina electrolysis is eliminated as a cause of the decay.

The effort put into the gas investigation was a substantial one both in time and materials, but the results were uniformly null in character. The general conclusion reached, therefore, is that, over the test period of 250 hours involving a fall of some 70% of the initial total emission, the contribution to the decay by gas action is negligibly small. It follows that the decay in the diode type 6D15 during this test period is primarily due to loss of excess barium by core solution and/or by evaporation.

All these experiments were carried out at a cathode temperature of 1 020°K.

(4.3) Absorption of Excess Barium by the Platinum Core

(4.3.1) Basis of Experimental Approach.

If a standard type 6D15 platinum-core diode is made up without its barium–strontium-oxide matrix it can still function as a diode provided that the bare platinum cathode is raised to a sufficiently high temperature. Some typical results of such a measurement are shown in Fig. 9, where the tubes have been

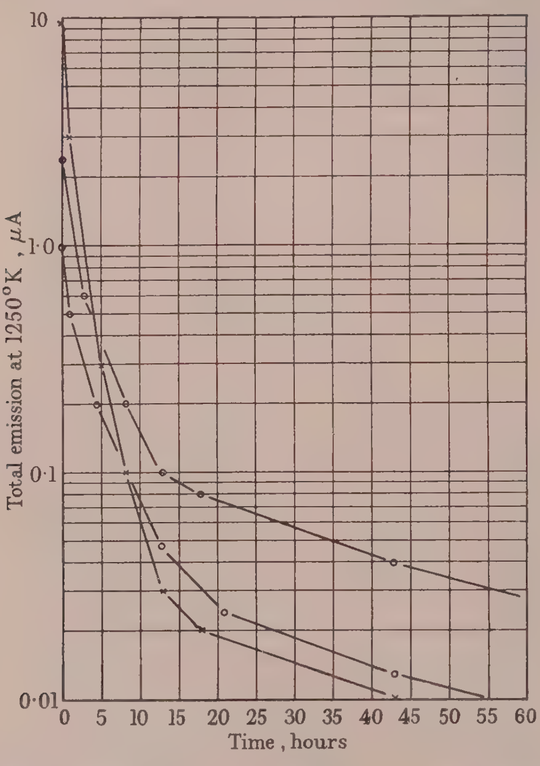


Fig. 9.—Emission/time characteristic with a bare uncontaminated platinum cathode.

run at a cathode temperature of 1 250°K and with 6·0 volts on the collector. The total emission is seen to rise to a few microamperes and then to fall steadily to a low value of around 0·01 μA. The relatively high initial value is doubtless due to contamination of the platinum by some low-work-function metal such as calcium which distils off as time proceeds. As the platinum surface becomes cleaner the work function approaches 5·3eV, and at 1 250°K this results in an emission of negligible proportion. If at this stage a trace of metallic barium is distilled on to the clean platinum surface, the work function will fall drastically and a very large increase in emission can be expected.

The comparison of total emission from a clean and a bariu contaminated platinum cathode is the basis of the method used investigate absorption of barium in the platinum core of t type 6D15 diode. After a supposed absorption has taken pla the platinum-core oxide cathode is removed from its dio structure, the barium–strontium-oxide matrix is stripped by acid process, and the bare contaminated core is reassembled in a type 6D15 diode. The total emission at 1 250°K is th measured over a time period using a new uncontaminat platinum-core diode as control.

(4.3.2) Experimental Detail.

A typical example will now be described to make the expe mental technique clear. Suppose it is desired to investigate t barium content of the core metal after the 250-hour decay. batch of six samples with a mean total emission of around 6m is run under zero-load conditions for 250 hours during which t mean total emission falls to 2·0mA. The cathodes are remove from the diode structure, and the barium–strontium-oxide coati is stripped by the following acid treatment:

Acid: Analytically pure hydrochloric acid diluted with equ volume of distilled water.

Time: 10min.

This treatment is repeated to give three immersions in al with renewal of the acid at each treatment. The cathodes a finally washed in distilled water.

The bare stripped cathodes are then reassembled into type 6D1 diodes and given the standard pump and process schedules f these valves. The batch is now ready for measurement where the cathodes are maintained at 1 250°K and the total emissio of each sample is measured by the application of 6 volts t the collector at regular intervals. The measurement is co tinued until the total emission has fallen to some low valu The required characteristic is the mean total emission plotte against time. A parallel control experiment is finally undertake using new platinum cores in the processed state attained pri to the normal barium–strontium-carbonate spraying operatio A typical result is shown in Fig. 10, where curve (*a*) represents th clean core and curve (*b*) the contaminated core after the 250-hou decay operation.

(4.3.3) Interpretation of Results.

It is assumed that excess barium has diffused into the platinu core from the overlay of the activated barium–strontium-oxid matrix. The acid stripping process will remove the matrix an destroy the excess barium lying exposed on the actual platinu surface, but it will not affect the platinum itself or the bariu lying beneath its surface. After reassembly and application o a temperature of 1 250°K to the cathode the interior barium wi start diffusing through to the platinum surface, and the tota emission will accordingly rise. This rise will continue until th surface concentration comes into equilibrium with the interio concentration. Once on the surface, however, the barium i subject to evporation, and this will lead to a progressive loss o barium from the core metal as a whole. The total emissio therefore starts to fall, and this trend continues until it reache the level appropriate to pure platinum.

This hypothetical picture of the total-emission/time charac teristic can be taken a step further. Suppose the core contain n atoms of barium and the average time that each atom reside on the cathode surface is t; then during surface residence th atom is responsible for passing an element of current i and a element of electrical charge it. Since all barium atoms mus eventually pass through the cathode surface, the total quantity Q, of electricity liberated from the surface is

$$Q = nit$$

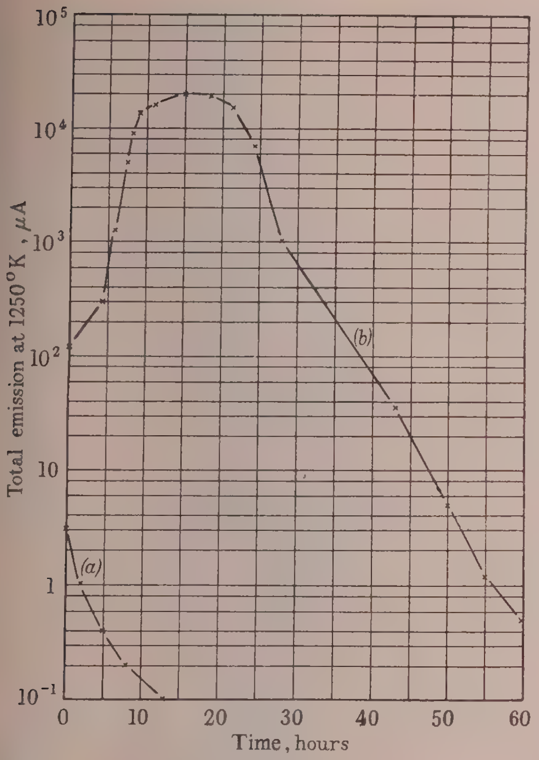


Fig. 10.—Illustrating effect of barium absorption by core metal.

At constant cathode temperature it is reasonable to assume that *it* is constant.

Therefore $Q \propto n$

This relation implies that the integrated area under the curve (*b*) of Fig. 10 is in direct proportion to the barium content of the core.

(4.3.4) Factors affecting the Degree of Barium Contamination of the Core.

The characteristics of Fig. 11 illustrate an attempt to gain some understanding of the factors that affect the degree of core contamination. Each curve represents the mean of six samples, and

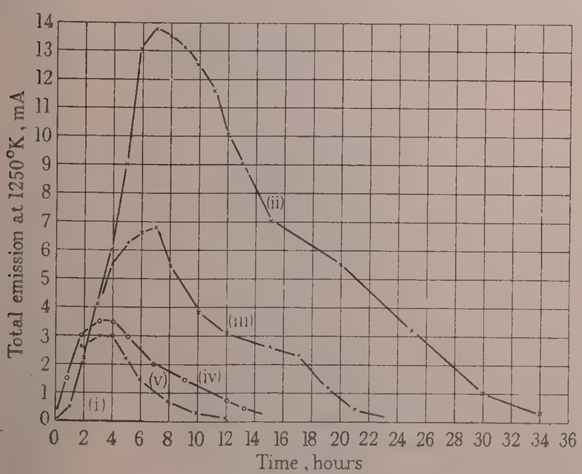


Fig. 11.—Conditions affecting degree of core contamination.

the determinations have been carried out under the conditions described in Section 4.3.2. The circumstances appropriate to each curve are listed in Table 3, where

V_H = Operating heater voltage during oxide-cathode decay period, volts.
t = Time of operation at voltage V_H, hours.
I_{T0} = Mean initial total emission at beginning of oxide-cathode decay, mA.
I_{T1} = Mean total emission after time t, mA.
$\Delta I_T = I_{T0} - I_{T1}$.
Q = Area enclosed by curves, mAh.

The whole series was undertaken with the oxide-cathode under zero-load condition. The bare platinum cores were measured for the function Q at 1 250°K.

Table 3

Curve No.	V_H	t	I_{T0}	I_{T1}	ΔI_T	Q	Remarks
	volts	hours	mA	mA	mA	mAh	
(i)	—	—	—	—	—	0·008	Pure platinum
(ii)	6·0	250	10·7	1·6	9·1	193	Contaminated platinum
(iii)	6·0	250	2·0	0·9	1·1	70	Contaminated platinum
(iv)	6·0	0	6·0	—	—	24·9	Contaminated platinum
(v)	3·75*	250	7·2	7·0	0·2	17·4	Contaminated platinum

* Corresponding to a cathode temperature of 850°K.

From the information given in Table 3 the following tentative conclusions are drawn:

(*a*) *Core Absorption.*—From curves (i) and (ii) it is concluded that a powerful absorption of barium into platinum has occurred during the activation and operation of the cathode.

(*b*) *Influence of* ΔI_T.—From curves (ii) and (iii) it is concluded that the quantity of barium absorbed by the core metal is a function of the range over which the total emission has decayed during the absorption period.

(*c*) *Zero Time Absorption.*—The presence of barium in the core metal [curve (iv)] immediately after manufacture shows that penetration occurs during this phase.

(*d*) *Influence of Temperature.*—Curve (v) shows that little barium passes from matrix to core at a temperature of 850°K, and this accords well with the observed lack of emission decay at temperatures below 800°K (see Section 3.3). Mobility of barium is clearly necessary both for core absorption and emission decay.

(4.4) Loss of Excess Barium by Evaporation

(4.4.1) Evidence of Loss.

The main fall of emission in the platinum-core diode type 6D15 over the first few hundred hours of life has now been traced to an actual physical departure of barium from the matrix. Of the two possible modes of egress it has been shown that core solution is a powerful contributor, and it remains to be seen whether evaporation has an effect of comparable magnitude.

That barium is evaporated from oxide cathodes is common experience and was perhaps first demonstrated in the classical experiments of Becker.[6] With active nickel-core cathodes the evaporation is so rapid that at 1 250°K a black stain of metallic barium is deposited on the valve envelope in 20 or so hours, and can be recognized by a sodium rhodizonate test. No such powerful reaction is observed with a pure platinum-core cathode.

and the act of evaporation must be demonstrated in more subtle fashion.

An experiment carried out by C. B. Johnson[7] at Dollis Hill will be used to demonstrate the evaporation and at the same time to introduce a physical phenomenon which is to be considered in the next Subsection. In this experiment two groups (50 each) of platinum-core pentodes type 6P12 were prepared under identical conditions and run on the life-test racks at the same cathode temperature but at different anode dissipations. In group A the anode temperature was 350°C, and in 5 000 hours this group had shown a significant growth of anode-control-grid capacitance. This capacitance growth is due to the deposition of barium films on glass envelope and mica. Group B runs with the anode at 240°C, and in the same period shows no sign of capacitance growth. A further point of interest is that the mean total emission of group A is about 50% higher than the mean value of group B. It is proposed by Johnson* to explain these events in the following terms. Barium is evaporated from the cathode in both groups, and this barium falls on the nickel anode plates. In group B with its cooler anode the barium settles and remains on the anode surface. In group A the barium falls on to a hotter anode surface, re-evaporates and passes back either to the cathode or to the insulating surfaces of the valve. This secondary evaporation gives rise to the higher total emission of group A and also to its growth of capacitance.

The hypothesis is neatly confirmed by placing the aged group B under conditions of group A—a very large but transient rise of total emission results immediately and the group ultimately increases in capacitance.

(4.4.2) **Rate of Evaporation Loss Relative to Rate of Core Absorption.**

The experiment described in this Section was based on the idea that barium is re-evaporated from a nickel anode maintained above 320°C. If the hot anode or reflector completely encloses the cathode, barium will be returned to the cathode as fast as it leaves and no resultant surface loss of barium will occur.

To develop this idea a type 6D15 diode was surrounded with a nickel shield running at a temperature of 345°C, and a decay curve was taken over the first 250 hours of life. The results are given in Fig. 12 as mean curves for a control batch and a shielded batch. Preliminary inspection shows that the rate and form of the decay is practically unaltered by shielding, but the total emission is everywhere increased in the shielded case. The reason for this upward shift of the characteristic has been examined and the following explanation is offered. The total emission, measured at 400 mW, has an ambient temperature coefficient of $+0{\cdot}6\%$ per deg C. The casing of the ordinary valve at 400 mW has an ambient temperature of around 22°C, corresponding to the envelope temperature; the shielded valve running at 400 mW has a shield temperature of 56°C. There is therefore a difference of 34°C in ambient temperature between the two cases, and this leads to an error of $20{\cdot}8\%$ in the total emission level. The temperature correction has been applied to the shielded case, and this brings them both into approximate coincidence.

Some thought has been given to the influence of this shield on the temperature of the cathode during the decay period. For a common heater voltage of 6·0 volts the shielded cathode will run a little hotter than the standard structure. This does not matter in respect of evaporation, since any increase in rate of departure is compensated by an increase in rate of return of barium. With regard to core diffusion, however, the decay rate should be increased, but there is no evidence in Fig. 12 to show that this is happening. It is possible, of course, that the increase in core-metal temperature is too small to show a detectable increase in diffusion rate, and this is confirmed to some extent by direct measurement of the tungsten-heater temperature in the two cases—at the same heater voltage no difference in heater temperatures could be detected.

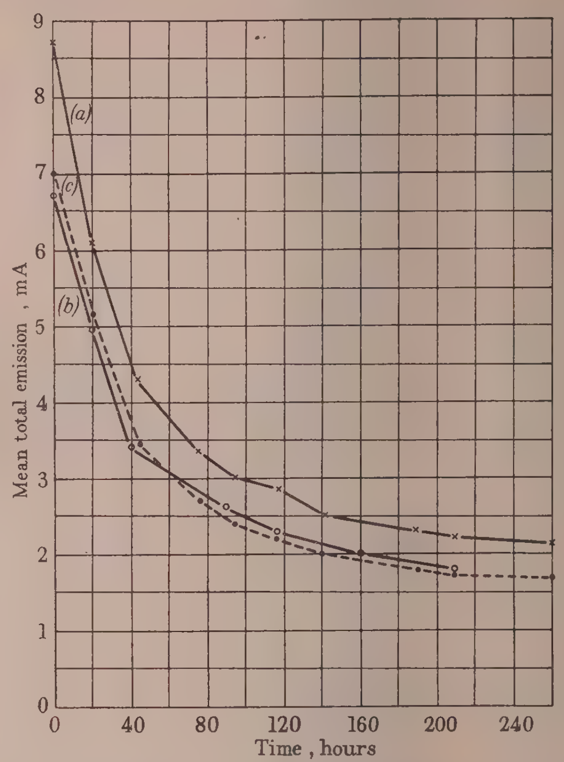


Fig. 12.—Influence of evaporation on initial decay characteristics.
(a) Shielded case. (b) Normal type 6D15. (c) Temperature correction.

In arranging the experiment it was found impracticable to enclose the cathode completely by the hot reflector, and end-gaps had to be accepted. The ratio of gap area to total area within the box was approximately 1 : 15, and it is assumed that the rate of evaporation loss from the system as a whole is reduced by this factor. The experiment is a slight one, but it gives support to the view that evaporation loss is not the predominant factor during the first few hundred hours of the decay.

(5) DECAY IN A PLATINUM-CORE CATHODE UNDER LOAD CONDITION

(5.1) Short-Term Decay

So far, all the work described has been carried out under zero-load condition, so that the decay effects can be examined without the confusing presence of a possible electrolytic reactivation by cathode current. The influence of a cathode load will now be studied, and Fig. 13 shows the relative influence of different levels of cathode current over a period of 900 hours. The curves are for mean values of ten samples in each group with all valves selected for an initial total emission of around 6 mA.

The drastic initial decay from 6 to 3 mA is seen to be hardly affected by the cathode load, and over this phase, covering around 50 hours, the predominating influence is core absorption. Below a total emission of 3 mA, however, the influence of current

* Mr. Johnson was assisted in this explanation by some useful conversations with Dr. J. J. Lander of Bell Telephone Laboratories.

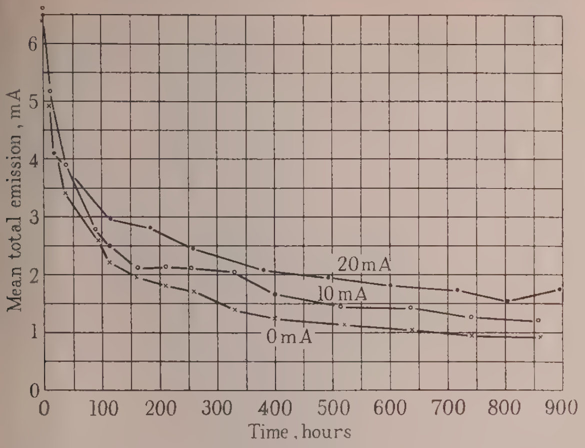


Fig. 13.—Influence of cathode current on decay characteristics.

becomes apparent, and the decay rate is inhibited by an amount which is some function of the current magnitude. It is hoped in the present Section to go some little way towards proving that a cathode current results in the continuous production of barium within the matrix, and this hypothesis will be used as an explanation of the influence of current magnitude on the decay rate.

(5.2) Long-Term Decay Characteristics

Sufficient has now been said to attempt a prediction of the long-term behaviour of a gas-free platinum-core-cathode valve running under load condition. Three factors have to be taken into consideration, namely core absorption, evaporation, and electrolytic production of barium, and the resultant of these actions will determine the total-emission level at any time. In the initial stage the predominant influence is core absorption, and this will shape the total-emission/time characteristic for the first hundred hours. As soon as the barium concentrations in core and matrix approach equilibrium, the influence of core absorption becomes negligible, and the two factors of evaporation and electrolysis remain to settle the total-emission level. Depending on evaporation rate and current the barium concentration in the matrix will approach a constant value, and the total-emission/time characteristic will become parallel to the time axis. This picture is, of course, essentially hypothetical, depending on the reality of the supposed electrolytic generation of barium.

Two actual cases of decay will be examined. The first concerns a group of platinum-core pentodes prepared with all known precautions against presence of residual gas, and the second concerns a similar group B in which residual gas is present. The two cases are shown in Fig. 14, and the gas-free group is seen to conform, so far as the test goes over 8 000 hours, with the picture presented above; i.e. a rapid initial fall is followed by a flattening of the characteristic to an approximately constant value.

The second case shows the behaviour of a group suffering the effects of residual gas generated from an imperfectly processed control grid. It has been shown elsewhere that the residual gas pressure in such circumstances is a maximum at zero time and decays in a roughly exponential manner. The rapid initial decay, therefore, is due primarily to gas deactivation of the matrix. As the pressure falls the destructive action becomes progressively slower until it comes into equality with the barium-producing electrolytic mechanism. Thereafter the electrolytic action is in the ascendancy, and the matrix recovers towards its equilibrium value.

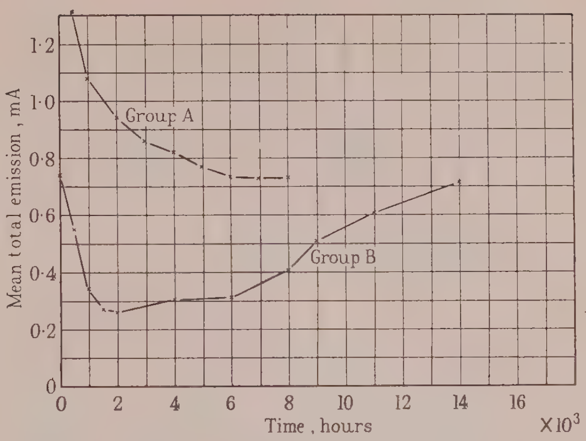


Fig. 14.—Effect of gas attack on emission of groups of working valves.

(5.3) Influence of Cathode Current on Storage of Barium in a Platinum Core

The technique described in Section 4.3 for examining the absorption of barium in the platinum core provides a useful tool for investigating the effect of cathode current on barium production in the matrix. If a current flow results in continuous barium production, it is reasonable to assume that a proportion of this barium finds its way into the core. In this Section a direct comparison is made between the barium content of two cores—one which has served its life under a matrix carrying current and the other under one carrying zero current.

The practical details of the measurements are similar to that described in Section 4.3.2, and each curve in Fig. 15 is the mean

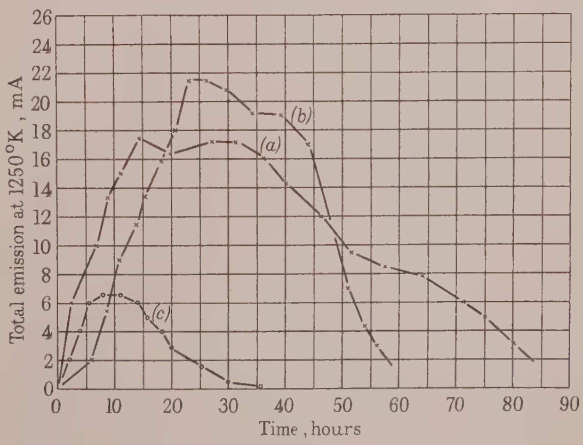


Fig. 15.—Influence of cathode current on barium storage in platinum core.

of six samples. The three cases reported were not set up deliberately for the present purpose but were withdrawn from the life-test racks as suitable subjects for investigation. Details of the life conditioning are set out in Table 4, together with the mean values of Q for each case. The bare-cathode emission/time curves from which the Q functions are derived are shown in Fig. 15.

The 6T15 type is identical in form to the type 6D15 diode, except that it is fitted with a third electrode and was run as a 100-volt triode during the life conditioning.

Table 4

Curve	V_H	t	I_c	Q	Valve type
	volts	hours	mA	mAh	
(a)	6·0	8 000	10	910	6D15
(b)	6·0	3 000	4	735	6T15
(c)	6·0	12 000	0	55	6D15

V_H = Heater voltage during life conditioning.
t = Time of life conditioning.
I_c = Cathode current during life condition.
Q = Quantity of electricity obtained from bare cathodes at 1 250° K.

The results appear to have considerable significance. The zero-load diode has accumulated no barium in its core and seems to have depended on its initial stocks—a result which shows that the reducing action of platinum on the barium–strontium-oxide matrix is of negligible importance. Both the current-carrying cathodes give unmistakable signs of substantial barium accumulation, and this can only have resulted from the action of cathode current.

(5.4) Probability of Dispensation of Barium from Platinum Core to Matrix during Valve Life

The experiments described in Section 5.3 suggest that core dispensation of barium to the matrix may exert a beneficial influence during the life of a platinum-core valve. If the matrix suffers a sharp fall of total emission following a transient gas attack, it can expect a flow of barium from the core to assist its reactivation.

This possibility has been examined experimentally. A batch of six valves was run under current load condition for 3 000 hours to stock the cores with barium. The cathodes were then stripped of their barium–strontium-oxide matrix in the usual way and remounted in the bare state as type 6D15 diodes. After increasing the emission to a maximum at 1 250° K (see Fig. 10) the batch was placed on long-term life test at the conventional cathode temperature of 1 020° K, and the total emission at this temperature was observed of a period over 6 000 hours. Results are plotted in Fig. 16 and indicate that after 6 000 hours the cathodes were still dispensing an appreciable supply of barium to the cathode surface. The emission from a pure platinum cathode at 1 020° K would be at least 1 000 times lower than the value reached at 6 000 hours on the contaminated cores. This dispensing action will be illustrated in greater detail in Section 5.7.

(5.5) Depth of Penetration of Barium into the Platinum Core

The depth of penetration of barium into the platinum core is of interest and has been investigated in two cases—one with a life history of current load and the other without load. The technique used is to cut the platinum surface to a known depth with aqua regia and then measure the core for function Q_c in the usual manner. In Table 5 results are set together with values of Q which which would be expected on the uncut surface.

The depth of cut d represents the thickness of the outer platinum surface removed, expressed as a percentage of the original platinum-core thickness.

Table 5

Batch No.	Depth of cut, d	Q_c	Q	Ageing condition
		mAh	mAh	
1	10%	0·0062	100	$I_c = 0$
2	1%	0·0087	100	$I_c = 0$
3	1%	0·0822	800	I_c = 4mA for 3 000 hours

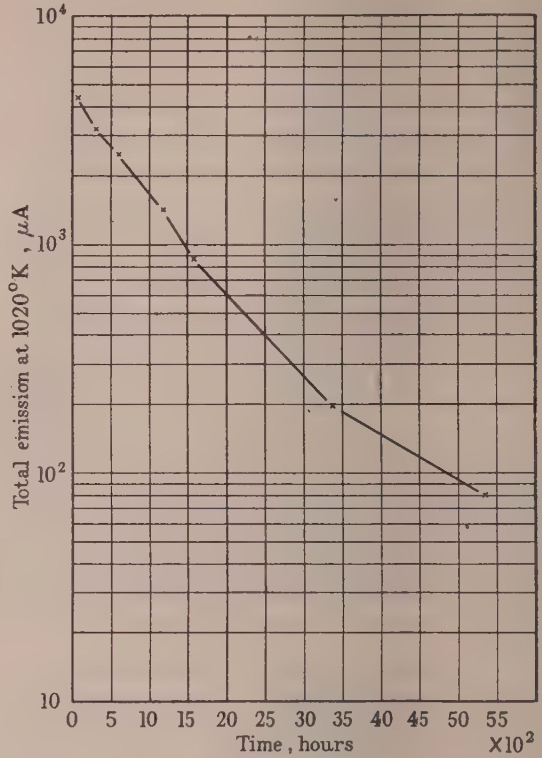


Fig. 16.—Dispensation of barium from a bare platinum core.

Batches 1 and 2 were zero-load conditioned and show that barium penetration is largely limited to about 1% of the platinum depth. Batch 3 shows a slight tendency to increased penetration under the impulse of cathode current.

(5.6) Influence of Cathode Current on Recovery of a Platinum-Core Oxide Cathode from Oxygen Attack

Useful information can be obtained by observing the behaviour of an oxide cathode under oxygen attack, and this general line of enquiry has received considerable attention at the Post Office Research Station. The example quoted in this Section is typical, and shows the influence of cathode current on recovery from an attack. Experimental procedure takes the following form. A platinum-core type 6D15 diode is processed, and its space-charge characteristic is measured on a high-grade bench pump using silicone oil as its working fluid. After processing the valve with its cathode at a temperature of 1 020° K, it is subjected to pure-oxygen poisoning at a pressure of 1×10^{-4}mm Hg for 16 hours. The gas supply is then cut off, the residue is pumped away to a pressure of 5×10^{-6}mm Hg, the getters are flashed, and the valve is sealed from the pump ready for experimental examination.

The cathode is first run under zero-load conditions at a cathode temperature of 1 020° K, and the space-charge condition is explored at intervals by a low-voltage pulse method over a period of 40 min. At the end of this period the voltage is applied continuously to the collector for another 40 min. The results are set out in Fig. 17 in the form of a percentage recovery with time of the initial space-charge current—the initial value being that measured before the application of the oxygen attack. The curve shows no recovery during the zero-load period, followed by a rapid and complete space-charge recovery under load.

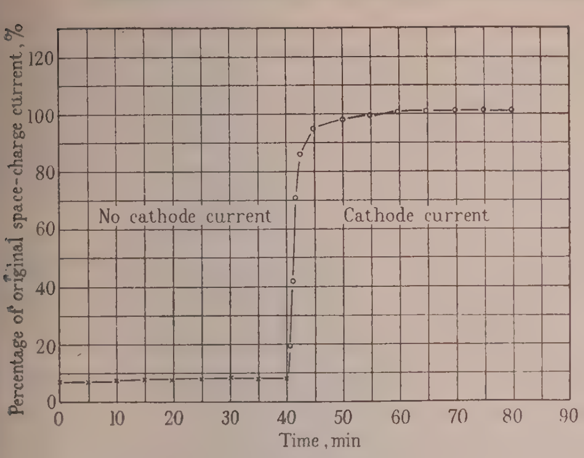


Fig. 17.—Influence of current on recovery of a platinum-core cathode.

This general type of poisoning-reactivation experiment is capable of many variants, and two will be mentioned as examples.

(5.7) Influence of Core Dispensation of Barium on Recovery from Oxygen Attack

It has been shown in Section 4.3.4 that there is a considerable core absorption of barium during the actual processing of a platinum-core valve. This absorption represents a reserve which may assist the cathode in recovery from an oxygen attack. To demonstrate this action the experiment of Section 5.6 is repeated but with the following variation: the zero-load phase is extended from 40 min to 800 hours to allow time for the diffusion of barium from core to matrix. During this extended period at a cathode temperature of 1 020°K the space-charge condition of the cathode is examined periodically by a low-voltage pulse method; the results are set out in Fig. 18. It is

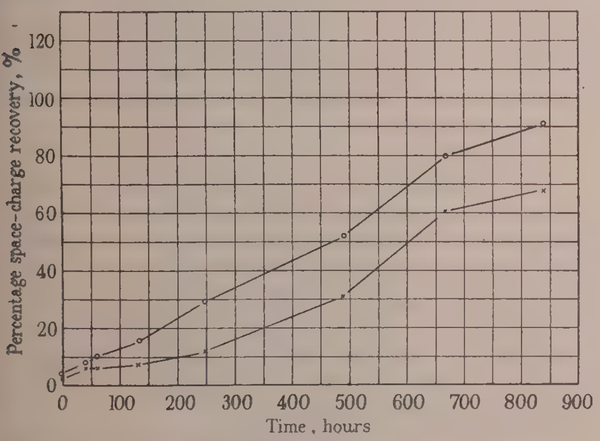


Fig. 18.—Influence of core dispensation on recovery from an oxygen attack.

apparent from the curves that considerable space-charge recovery has taken place without the aid of cathode current, but the rate of recovery is very slow. The phenomenon is assumed to be due to core dispensation.

(5.8) Nature of the Current-Dependent Activation

The passage of an electron current through the oxide-cathode sets up a potential gradient across the matrix. The hot barium–strontium-oxide ionic lattice is thus under electric stress, and it is assumed that this results in a straightforward electrolysis. There is some evidence to support the idea of such a Faradic action in the manner of the recovery of cathodes from oxygen attacks. It has, for example, been found that the recovery rate is very dependent on the level of current traversing the cathode. At a cathode temperature of 1 020°K and an oxygen attack of the type recorded in Section 5.6, recovery at a current of 30 mA takes about 30–40 min, whereas recovery at 3 mA takes 10–15 hours. The recovery is assumed to be complete when the normal-temperature total emission has reached 1 amp/cm^2. These cases seem to indicate a dependence of recovery on the quantity of electricity.

(5.9) Conclusions

The primary object of Section 5 has been a search for the existence of a mode of excess barium production based on cathode current. Three points of significance have emerged:

(*a*) The dependence of the total-emission level on current level.
(*b*) The accretion of barium by the platinum core under load condition.
(*c*) The dependence of emission recovery on cathode current after an oxygen attack.

On the whole the evidence is regarded as strongly supporting a theory of continuous, current-dependent activation of a cathode.

(6) HYPOTHETICAL EMISSION-LIFE MODEL OF A PLATINUM-CORE CATHODE

(6.1) Under Zero Load

Under zero load and freedom from gas attack the behaviour of the platinum-core valve appears relatively simple. In a well-activated cathode with high initial total emission, the emission decay over the first few hundred hours is dominated by the phenomenon of core absorption. As the barium concentrations of matrix and core approach equilibrium the rate of total-emission decay falls to a low value, and the phenomenon of evaporation becomes the prime compelling factor. During this long-term phase of evaporation loss, the barium-concentration gradient between matrix and core changes direction, and the core begins to dispense barium to the matrix. This dispensation has the effect of inhibiting the rate of decay due to evaporation. Despite core dispensation, however, the trend is towards loss of barium from the cathode system as a whole, and the decay tendency is regarded as continuing indefinitely until the matrix concentration comes into equilibrium with the partial pressure of barium within the tube.

(6.2) Under Load

The reaction of the cathode under a constant-current load is obtained by superimposing a constant reactivation effect on to the zero-load model. Decay is everywhere slowed up until it ceases altogether at the point in time when evaporation loss is balanced by electrolytic production of barium.

(6.3) Under Gas Attack

Experience shows that gas attack in valves may take two forms: one is finite in duration decaying in approximately exponential fashion, and the other is indefinitely continued. The first form is a thermal outgassing phenomenon, and the second results from electron bombardment of insulating parts.

The cathode reaction under load follows from an application of the principle of superposition. With the decaying type of attack the total emission falls below the equilibrium value of Section 6.2 and then slowly recovers to the normal equilibrium level as the gas pressure is withdrawn. Under a continuous

attack the total emission decays to a constant level below that of Section 6.2 and dependent on the relative magnitudes of evaporation rate, gas destruction rate and electrolytic reactivation rate.

(6.4) Comment

This life model of the platinum-core oxide cathode is based on the experimental evidence of previous Sections together with two further assumptions—that the electrolytic action is invariant with time and that the behaviour of the cathode towards oxygen is representative of all other gases found in the working valve envelope. The picture presented is no more than a working hypothesis which is in harmony with a reasonably wide range of observed experimental facts.

(7) CONCLUSIONS

Two main conclusions are drawn from Part 1 of the paper. First, the behaviour of the gas-free platinum-core oxide cathode can be defined in terms of core solution, evaporation, core dispensation and electrolytic regeneration. Secondly, the cathode is largely immune to the action of oxygen. Once the attack is withdrawn the necessary excess barium is regenerated by current action.

(8) APPENDICES

(8.1) Pump Processing Detail for Standard Diode Type 6D15

(*a*) Seal valve on to pump and evacuate to a pressure of 5×10^{-5} mm Hg.

(*b*) Bake for one hour at 400° C.

(*c*) Eddy-current heat collector. Three doses of 15 sec duration at 900° C. Degas getters.

(*d*) Decompose cathodes to the following schedule:

Time	Heater voltage
min	volts
1	4·0
1	6·0
1	8·0
1	10·0
4*	12·0

* Or until manifold pressure falls to 5×10^{-6} mm Hg.

(*e*) Repeat as for (*c*) with $V_H = 11$ volts.

(*f*) Collector bombardment:

Time	.. ..	5 min
Heater voltage	.. ..	11·0 volts
Collector voltage	.. ..	10·0 volts
Collector current	.. ..	100 mA

Flash getters and seal off pump.

(*g*) Further activation:

Time	Heater voltage	Collector voltage
min	volts	volts
½	11·0	10·0
5	7·0	10·0

Allow valves to cool and measure total emission at 400 mW. No further treatment is given to the diode after this stage.

(8.2) A Few Details on Piece-Part Pre-Processing

(*a*) *Glassware.* Conventional treatment.

(*b*) *Mica.* Vacuum stoved at 550° C for 16 hours, and stored in vacuum.

(*c*) *Metal piece-parts.* Degreased and then hydrogen stoved for 5 hours at 950° C. Special precautions are taken to maintain hydrogen purity.

(*d*) *Carbonates.* The carbonate sludge used throughout the work is a conventional co-precipitated one of barium–strontium carbonates in equimolecular proportion. The binder is a nitrocellulose one, and the sludge is applied in the normal way by spraying to a known thickness and density.

Part 2. EMISSION BEHAVIOUR OF AN OXIDE CATHODE ON PASSIVE AND ACTIVE NICKEL CORES

(9) INTRODUCTION

The barium-oxide-on-platinum system has been found to be essentially simple. The core metal is chemically inert and enters into none of the basic cathode reactions apart from its absorption effect. This absorption lends to the core its one positive function—that of acting as a "bank" of excess barium.

The nickel-core cathodes are more complex in behaviour, since the base metal itself and its useful contaminants are chemically active and play an influential part in the emission life of the cathode. It will be found helpful to study these chemical actions in isolation, and a start will therefore be made by examining the pure or passive nickel case. This is very similar to platinum in behaviour but with two differences: the absorption effect is largely absent and the core metal is strongly reactive to oxidizing residual gases at the working temperature. Once the pure nickel-core behaviour is appreciated, the more complex case of the active core can be approached with greater equanimity.

(10) PASSIVE NICKEL-CORE BEHAVIOUR

(10.1) Core Metal

Very pure nickel is difficult to obtain, and in the present work a supply of nickel has been used that, although much purer than the conventional active-core metal, is still capable of some reducing action. Two supply sources were considered, one from an electrolytic method and the other from a powder-metallurgy carbonyl process. Quantitative spectrographic analysis of the two sources gave very similar results, and the data quoted below refer to the electrolytic type:

Silicon	0·003%
Magnesium ..	0·001–0·0005%

These figures must be treated with reserve, but they indicate that the activity of the passive metal is between 10^{-1} and 10^{-2} of that of the conventional active cores.

It will be clear from these considerations that the passive nickel is pure only in the sense that it has a much lower level of activity than the normal core. Experience has shown, however, that this difference is a useful experimental device and leads to a contrasting behaviour from which certain chemical actions can be reasonably inferred.

The relative levels of gross reducing power can be better appreciated from a simple experiment. Two batches of type 6D15 diodes are prepared, one with passive and the other with active nickel cores. After the usual processing the two batches are run at 1 250° K for 50 hours. At the end of the run the active-core batch has developed an intense black stain of metallic barium on the glass envelope opposite the oxide-cathode; the stain in the passive-core batch is only just visible on careful inspection. That the stains in both cases are primarily barium is proved by opening the envelopes to air, whereupon the stains instantly disappear.

(10.2) **Emission Decay under Zero-Load Condition**

The mean zero-load decay characteristic of a group of type 6D15 diodes fitted with passive nickel cores is shown in Fig. 19,

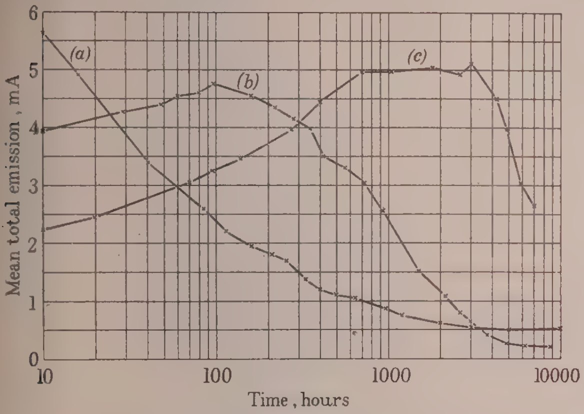


Fig. 19.—Zero-load decay characteristics.
(a) Platinum. (c) Active nickel.
(b) Passive nickel.

together with the curve for platinum as a reference (the curve for active nickel should be ignored at this stage). The behaviour of the two cathode types is strikingly different, and may be summarized in the following way:

(a) Over the first 1 000 hours the mean emission in the passive-nickel case is much superior to that in the platinum case.

(b) After about 4 000 hours the characteristics converge, and thereafter the order of emission merit is reversed.

(c) With the passive nickel core the emission rises to a maximum over the first hundred hours, whilst with the platinum core it falls continuously.

These differences in behaviour must be due to one or both of the following causes: the presence of the low residual level of reducing power in the passive core, or differences in the basic properties of platinum and nickel. The first of these alternatives will be examined in this Section and the second will be considered in Section 10.4.

If the difference in behaviour is due to the presence of the residual reducing power of the passive nickel core, the degree of difference should be lessened by a lowering of the reducing power. The idea was put to the test in the following manner: A group of ten passive nickel cores were mounted in type 6D15 structures and given the usual pump process. The bare cores were then run at a temperature of 1 500° K until a dense cloud of the core metal had evaporated on the glass envelope. This process caused a heavy loss of the volatile magnesium contaminant and resulted in a lowering of the gross reducing power of the core. The cores were then recovered, sprayed with carbonates and processed in the usual way as type 6D15 diodes. A test over 6 000 hours resulted in a mean characteristic which showed no significant difference to that shown in Fig. 19. It is concluded, therefore, that the basic differences between the two characteristics of Fig. 19 are not due to the presence of the residual reducing power of the passive nickel core.

(10.3) **Investigation into Relative Rates of Absorption of Metallic Barium by Passive Nickel and Platinum Cores**

The difference in behaviour shown in the zero-load decay characteristics of passive-nickel and platinum cores can obviously be explained by a difference in the capacity of the two metals to absorb barium. That such a difference exists and is indeed very large can be shown by the following experiment. Two batches of platinum and passive-nickel cores were mounted bare in envelopes and given the usual vacuum processing. Each core was then set at a temperature of 1 000° K, and about 2·5 mg of barium metal were distilled on to its hot surface from a getter. This distillation was carried out with the hand applicator of an eddy-current-heating machine in such a way that the period of distillation covered about 30 sec. The cores from both batches were then recovered by opening the envelopes under fresh distilled water, dried and remounted as bare-core type 6D15 diodes. After vacuum processing both batches were run at a cathode temperature of 1 250° K and a record was taken of the respective Q figures in milliampere-hours. The mean results for four such determinations are given in Fig. 20.

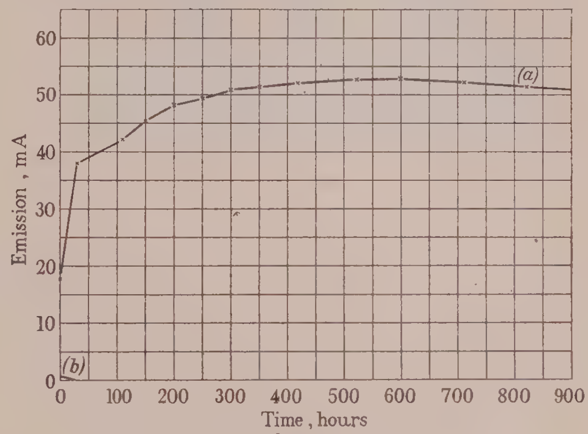


Fig. 20.—Characteristics of Q figures.
(a) Platinum.
(b) Nickel.

Comparison of the two characteristics shows that the platinum batch is immensely more active thermionically than the nickel batch. The Q figures derived from the characteristics are as follows:

$$Q_{Pt} > 40\,000\,\mathrm{mAh}$$
$$Q_{Ni} = 0{\cdot}78\,\mathrm{mAh}$$
$$Q_{Pt}/Q_{Ni} > 50\,000$$

If it is assumed that the Q figures indicate the relative absorption rates, it seems that at 1 000° K platinum absorbs barium at least ten thousand times faster than does nickel.

(10.4) **Interpretation of Results**

An attempt will be made to explain the differences in the form of the zero-load decay characteristics of oxide cathodes on platinum and passive nickel cores. For the purposes of analysis it will be assumed that the two systems are mechanically identical, have no reducing power, run at the same temperature under gas-free conditions, and that platinum alone has the power to absorb barium. Such a system has no means of generating barium and can only lose the metal by evaporation. It will further be assumed that the whole of this loss is by way of the outer surface of the oxide matrix. The rate of barium loss from this surface is directly proportional to the concentration n' of barium atoms lying on the surface, and from the known proportionality between surface emission and matrix conductivity, the loss rate can be related to total emission by

$$\text{Rate of barium loss} \propto n' \propto (\text{total emission})^2$$

The total barium loss over a period of time is thus a function of the integral of the (total emission)²/time characteristic, and the losses of the two systems can be directly compared since their outer matrix surfaces are identical in all respects except concentration of barium. A derived (total emission)²/time characteristic is given in Fig. 21, and integration of such curves over a period of 3 500 hours gives the following results:

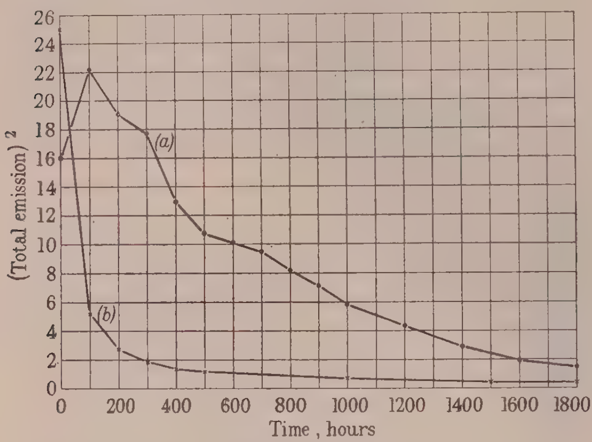


Fig. 21.—Derived (total emission)²/time characteristics.
(a) Pure nickel.
(b) Platinum.

$$\text{Nickel} \qquad \int_0^{3\,500} (I_{T1})^2 dt = 660 \text{ units}$$

$$\text{Platinum system} \qquad \int_0^{3\,500} (I_{T1})^2 dt = 143 \text{ units}$$

$$\text{Barium balance} \qquad = 517 \text{ units}$$

The period of 3 500 hours has been chosen since it coincides with the cross-over point (Fig. 19) of the two characteristics where both matrices must have the same barium concentration. Assuming equality of barium concentration at zero time, it seems that at 3 500 hours the platinum system still "banks" in its core metal a balance of 517 units, or 80% of its original stock, whereas the nickel-system stock is approaching exhaustion. The probability of further life after 3 500 hours thus appears more likely for the platinum than for the nickel system. The ideal core-matrix system would therefore seem to be one capable of transferring almost the whole of its zero-time barium stock to the core and then feeding the metal back into the matrix at a rate just sufficient to maintain the total emission at an adequate working level. The level accepted as adequate for the 2-watt cathode submarine repeater valve is about 150 μA. This analysis covers points (a) and (b) of Section 10.2.

The third significant difference in the characteristics of Fig. 19 is the relative behaviour over the first 100 hours of test: with the platinum system the emission falls continuously, whereas with the nickel system it rises to a maximum at about 100 hours and then starts to decay. This difference is of interest, and an attempt will be made to explain it on the following lines. Imagine an oxide-cathode system in which the barium–strontium-oxide matrix is stoichiometric. Owing to the absence of excess barium the work function of the system will be high and the emission very low. If barium is now admitted to the matrix the emission will rise and the work function will fall progressively to a low value of about 1·1 eV. As injection of barium continues the work-function trend is reversed, and it rises asymptotically to a limiting value of 2·4 eV, as the ratio of barium to barium–strontium-oxide becomes infinitely great.

Minimum work function or maximum emission is then identified with an optimum value of the ratio of barium to barium–strontium-oxide. The rising characteristic of the nickel system is explained in the following terms. Up to 100 hours the ratio of barium to barium–strontium-oxide is greater than the optimum value, and the total emission rises because the matrix is losing barium by evaporation. At 100 hours the ratio is at an optimum, and thereafter it falls together with the total emission. In the platinum case it is assumed that the rapid absorption of barium into the core metal during actual manufacture leaves the matrix with the ratio of barium to barium–strontium-oxide at or less than the optimum value at the beginning of the test life.

This rather slight hypothesis has some little experimental backing. Examination of a number of type 6D15 oxide cathodes on platinum cores showed a proportion which did in fact rise to a maximum total emission before starting the decay. The time at which the maximum occurred was, however, always less than 5 hours. In a more deliberate experiment a type 6D15 diode was processed in the presence of a barium generator of the batalum type. After the usual vacuum treatment, which included thorough outgassing of the batalum element, some barium metal was distilled on to the cathode at 1 020° K. The emission fell during the period of distillation and then slowly recovered to its previous value as the condensed barium on its surface dispersed by evaporation. This effect may be due either to a temporary increase in work function or to an increase in thermal emissivity. The result is inconclusive but not destructive to the hypothesis.

(11) OXYGEN ATTACK ON THE PASSIVE NICKEL CORE

(11.1) Comparative Behaviour with Platinum Core

The first of the two major differences between the core metals lies in their relative ability to absorb barium; the second concerns their reaction to oxygen attack. Platinum is unaffected, but nickel is powerfully oxidized at 1 020° K, and this leads to a remarkable contrast in behaviour of the two cathode systems. A simple experiment will illustrate the point. Two type 6D15 diodes—one fitted with platinum and the other with passive-nickel-core oxide cathodes—were mounted on a bench pump, processed and then subjected to 16 hours of oxygen poisoning at a pressure of 1×10^{-4} mm Hg and a cathode temperature of 1 020° K. After pumping away the gas, the pair were sealed from the pump, getters were flashed and the cathodes were run for 30 min at 1 020° K with 8 volts applied to the collectors. The voltage/current characteristics before the oxygen attack and after recovery treatment are shown in Fig. 22. The platinum-core case has completely recovered whereas the nickel case shows no sign of revival whatsoever. On further running of the nickel cathode at 1 250° K no further improvement occurred. This radical difference in behaviour can only be due to core oxidation in the nickel case.

(11.2) A Reaction between Nickel Oxide and Barium Oxide

It is assumed in the experiment of Section 11.1 that an interface layer of nickel oxide forms during the oxygen attack and that this layer prevents a subsequent reactivation. The existence of the layer seems to be proved in the following experiment. Two sets of type 6D15 diodes with passive nickel cores were prepared,

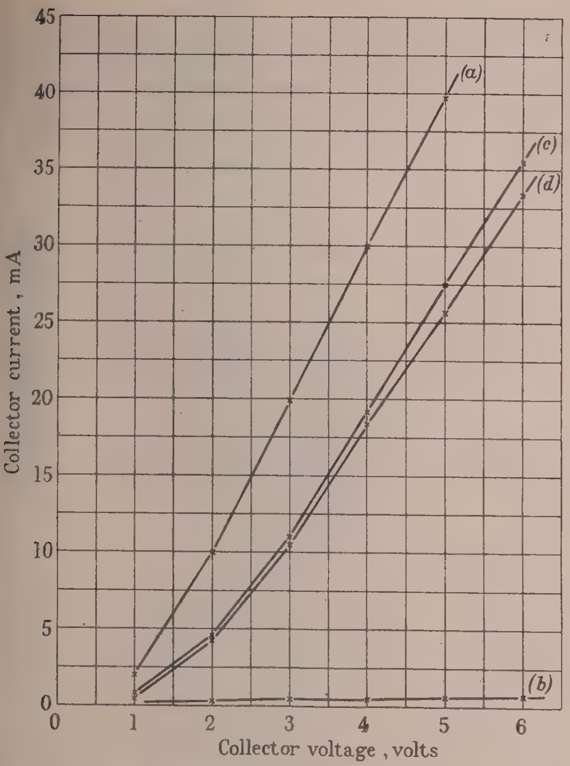


Fig. 22.—Voltage/current characteristics before and after oxygen attack.

(*a*) Nickel before attack. (*c*) Platinum before attack.
(*b*) Nickel after attack. (*d*) Platinum after attack.

and after vacuum processing, one set was subjected to oxygen poisoning at a pressure of 1×10^{-4} mm Hg for 16 hours at 1 020°K. After pumping off the gas, getters were flashed and both groups were subjected to a steady increase of cathode temperature. The normal group showed no change of appearance of the oxide-cathode surface as the heater voltage was increased to 20 volts; the oxygen-poisoning group, however, gave a powerful reaction when the heater voltage reached about 15 volts. This reaction takes the form of a vigorous fusing of the matrix from the normal white granular structure to a purple-brown fluid. The cathode temperature at which the reaction is first observed is about 1 400°K. Typical samples from both groups are shown in alternate positions in Fig. 23.

The chemical reaction underlying the phenomenon is thought to be the formation of a fusible barium nickelite of the barium–nickel-oxide type, which is known to follow the interaction of barium oxide and nickel oxide at red heat. Two interesting points of practical importance to working valves arise from the question whether the nickel-oxide interface layer is stable with time and whether the nickelite reaction occurs at the normal working temperature of 1 020°K.

(11.3) Stability of the Nickel-Oxide Layer

A useful idea of the stability of the nickel-oxide layer with time at 1 020°K can be gained from the following experiment. Two batches of type 6D15 oxide-cathode diodes with passive nickel cores were vacuum-processed. One batch was held as a control, and the other was submitted to oxygen poisoning at a pressure of 1×10^{-4} mm Hg for 16 hours at 1 020°K. After pumping away the gas, getters were flashed and the valves were sealed from the pump. Both batches were then run under zero-

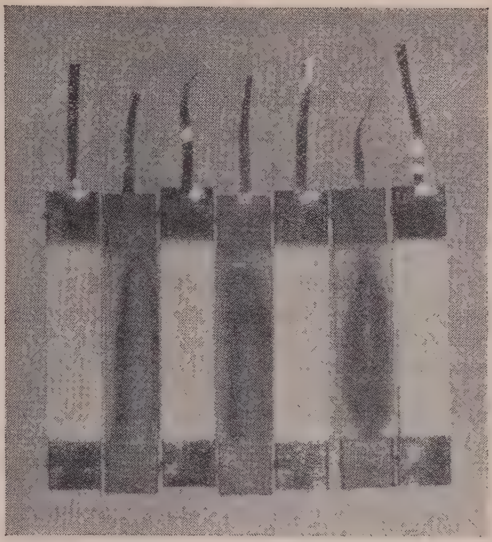


Fig. 23.—Reaction between barium oxide and nickel oxide.

load condition for a test period of 500 hours at 1 020°K. At the conclusion of the run both batches were subjected to a steadily increasing cathode temperature; the control batch was unaffected, but the oxygen-treated group showed the intense nickelite reaction at about 1 400°K. It is concluded that the nickel-oxide layer is relatively stable at 1 020°K. Microscopic examination of the cathode surfaces before the overheating treatment showed no differences between the batches, and this seems to show that prolonged heating at 1 020°K does not initiate the nickelite reaction.

(11.4) Reduction of the Layer by Hydrogen

That the nickel-oxide layer is readily reduced by hydrogen is shown by the following experiment. Three passive nickel-core type 6D15 diodes were subjected on the pump to an oxygen attack at a pressure of 1×10^{-4} mm Hg for 16 hours at 1 020°K. The gas was then pumped away and one valve was sealed off as a control sample. The two test cathodes were then set at a temperature of 1 020°K and subjected to the action of pure hydrogen at a pressure of 1×10^{-4} mm Hg with the collectors at 8 volts. Complete space-charge recovery was effected after 3 hours of this treatment. Attempts to recover the control sample after the firing of its getter were unsuccessful, and after five hours' treatment at 1 300°K it had regained only 10% of its original space-charge current.

(12) ACTIVE NICKEL-CORE BEHAVIOUR

It has been found useful to describe the behaviour of the passive nickel core as similar to the more basic platinum case, but with two differences; i.e. prevalence to core oxidation and an absence of core absorption. Following this precedent the case of the active nickel core will be based on the behaviour of the passive core, the one and only difference between the two cases being the introduction of powerful reducing agents capable both of direct chemical production of barium from the matrix and of reduction of the nickel-oxide layer. It will suffice, therefore, to limit work on the active core to a study of these reducing actions.

The active-core nickel used for experiment in this Section contained approximately 0·08% magnesium and 0·08% silicon, and was typical of the core metal used in industry. Both contami-

nants are capable of direct reduction of barium oxide on the following lines:

$$Mg + BaO = MgO + Ba$$
$$Si + 4BaO = Ba_2SiO_4 + 2Ba$$

and the by-product of the silicon reaction is the undesirable orthosilicate interface layer. Calculation of the gross reducing power of the alloy shows that a typical cathode core is capable of producing several hundred times the quantity of barium necessary for full activation of the matrix overlay.

(12.1) Reducing Power at 1 250°K

During manufacture, valve cathodes are taken up to a temperature of 1 200–1 300°K for a period ranging from 5 to 30 min, and the reducing power of the core under these conditions is of interest. The action can be demonstrated in the following manner: Two groups of type 6D15 diodes, one fitted with platinum and the other with active nickel cores, were processed and run at a cathode temperature of 1 250°K for 50 hours. The active group developed an intense black stain of metallic barium on the glass envelope whilst the platinum group showed no stain at all. The stain was proved to be barium by its reaction to a rhodizonate test and its instant disappearance in air.

(12.2) Reducing Power at 1 020°K

Reducing power at the conventional operating temperature of 1 020°K can be demonstrated as in Section 12.1. At the lower temperature the action is, of course, much slower, but the envelope stains of barium in valves of the active group are visible after 3 000 hours and fairly dense after 5 000 hours. No stains at all are visible in the control passive-core batch. Both batches are aged under zero-load condition, and must be processed to a sufficiently high degree of vacuum to avoid oxidation of the deposited films during the test run.

It is clear that reducing power is available at the conventional operating temperature and must be considered as a potent factor in maintaining emission in early life.

(12.3) Decay of Reducing Power during Life

(12.3.1) Probable Form of Decay.

The reducing power of an active nickel core is a finite quantity, and must suffer decay as it does useful work. To perform such work the reducing element must cross the core-matrix boundary and enter the matrix itself. If it is assumed that this boundary is freely traversed, the problem of decay is a relatively straightforward exercise in diffusion theory. Thus if R_0 is the reducing power at zero time, the decay factor will take the general form

$$R_t = R_0 \varepsilon^{-\lambda t}$$

The factor λ can be expressed as a half-life function, and this itself will be an exponental function of core temperature.

(12.3.2) Investigation of Decay in Working Valves.

So far as the author is aware there is no published work on the problem of decay of reducing power with life. This is not altogether surprising, since such an investigation would require deliberate planning and a wait of years before useful results could be obtained. The general lines on which such a test would be based would be the assembly of a large group of identical valves and their ageing in batches for times ranging from zero hours to 50 000 hours or more. The cores would then be recovered and examined for changes in concentration of the reducing agents.

No such deliberate experiment has been arranged at the Post Office Research Station, but by chance a small group of common pentode valves which fulfilled the general requirement were found in the Life-Test Section. These valves came from a single factory run and were all aged under similar conditions. The cathodes were recovered, stripped of the matrix overlay by 10% acetic acid and distilled-water washing, and then subjected to quantitative spectrographic analysis for magnesium. The results are given in Table 6 and the mean values in Fig. 24.

Table 6

Life, hours	0	4 000	36 500	67 000
Individual determinations, % magnesium	0·076 0·085 0·072 0·072 0·068 0·081	0·063 0·046 0·057 0·058 0·055 0·062	0·039 0·049	0·038 0·048 0·055 0·048 0·041 0·042
Mean, %	0·076	0·057	0·044	0·045

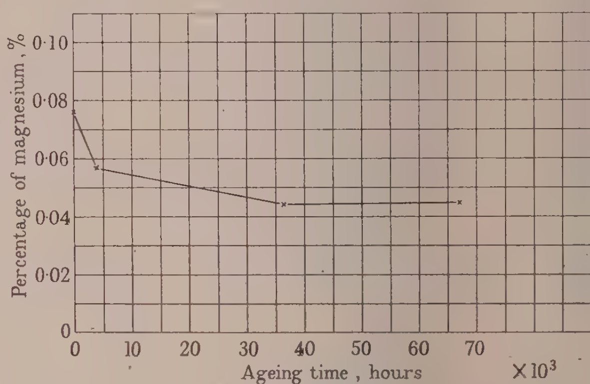


Fig. 24.—Decay of magnesium with working life.

Each of the figures in Table 6 represents the result obtained on one cathode core, and the individual values are set up to help assess the significance to be given to the mean results. The batch is a small one, but there seems to be evidence to show that the bulk of magnesium decay occurs within the first 10 000 hours of life at 1 020°K. The levelling-out of the characteristic at a magnesium concentration of 0·044% may be due to one or both of the following causes:

(*a*) The zero-time magnesium concentration of 0·076% may contain up to 0·044% in oxide form.

(*b*) The growth of the barium-orthosilicate interface layer may inhibit the passage of metallic magnesium from core to matrix after the first few thousand hours of life.

No attempt was made to determine the decay of silicon with time, since it was found difficult to separate the sample from the orthosilicate layer.

A second and rather different approach to the problem was made on three groups, each of six samples, which had aged for 0, 4 000 and 14 000 hours, respectively. The cathodes were recovered, stripped of matrix in the usual manner, resprayed with carbonates and remounted as type 6D15 diodes. All the valves were then pumped to a gas-free condition, the getters were fired and the cathodes were run at 1 250°K for 50 hours. This is a simple test of gross core activity, which is measured by the intensity of the barium-metal stain deposited on the glass envelope. To give the test some little quantitative significance the stain densities were matched against a standard photographic

stepped density wedge* having a span of 21 steps ranging from the barely visible (No. 1) to the almost black (No. 21). The mean intensity number for each group of six samples is plotted in Fig. 25 against the ageing time in hours of the cathodes at

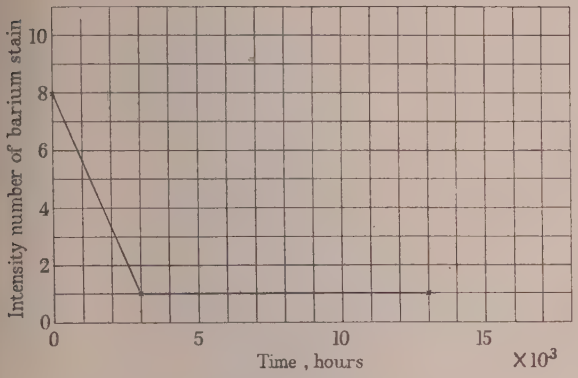


Fig. 25.—Stain-intensity-number/time characteristic.

1 020°K. The test is undeniably crude, but it has the merit of showing the trend of gross effective reducing power in striking fashion. The results are consistent with the spectrographic evidence, and show a rapid decline in the first few thousand hours.

nickel and active nickel. The supposedly significant chemical reactions are set out in Table 7, in which R represents the reducing element in the active-nickel case.

It has been shown that the nickel-oxide interface found in the passive-nickel case is relatively stable with time at 1 020°K and prevents the build-up of an electrolytic action which might lead to rapid reactivation of the cathode. That a similar nickel-oxide layer occurs with the active core can be taken for granted, but in this case means are available for its reduction. If platinum and active nickel cores are subjected to a common oxygen attack of limited duration, it can be anticipated that the platinum cathode will recover rapidly by electrolysis whilst the active case will react in more leisurely fashion depending on the rate of flow of reducing agent to the oxide interface. The rate of recovery of the active core will therefore be a function of temperature, and its completion will only be possible if there is at least a chemical equivalence between gross reducing power and the mass of the nickel-oxide interface layer. These predictions are substantiated by experiment, and Fig. 26 shows a typical recovery characteristic from an oxygen attack at a pressure of $5 \cdot 0 \times 10^{-5}$ mm Hg for 10 min at 1 020°K. After shutting off the gas the valves were allowed to pump down for 10 min and then sealed from the manifold. No getters were used in the tubes, and recovery was effected at 1 020°K with a voltage of 8 volts on the collectors. The recovery curves are expressed as percentages of the space-charge current given by each tube before the onset of poisoning.

Table 7

Platinum	Pure nickel	Active nickel	Phase
$2Ba + O_2 = 2BaO$	$2Ba + O_2 = 2BaO$	$2Ba + O_2 = 2BaO$	Poisoning
—	$2Ni + O_2 = 2NiO$	$2Ni + O_2 = 2NiO$	Poisoning
—	—	$NiO + R = RO + Ni$	Recovery
—	—	$BaO + R = RO + Ba$	Recovery

(12.3.3) **Comment.**

The evidence of Section 12.3.2 is slight and leads to little more than an impression of fairly rapid decay of conventional reducing power at 1 020°K. If a simple exponential form of decrement holds good, an estimated half-life of 10 000 hours at 1 020°K would lead to virtual extinction of reducing power at 50 000 hours.

(12.4) Zero-Load Emission Decay Characteristic

The mean zero-load decay characteristic at 1 020°K for a group of ten active nickel cores is shown in Fig. 19, together with the characteristics for passive nickel and platinum as a reference. The only known difference between the two nickel cases is in the concentration of reducing power. The two characteristics are similar in form, and the effect of increasing the reducing power is to displace the emission peak along the time axis by about 3 000 hours. This shift is thought to be due to excess barium production by the reducing agents delaying the approach of the matrix to the optimum ratio of barium to barium–strontium-oxide. If this view is accepted, it must be extended to include the cause of rapid decay of total emission after 3 000 hours as being due largely to the decay of reducing power.

(13) OXYGEN ATTACK ON THE ACTIVE NICKEL CORE

(13.1) Comparative Behaviour of Platinum and Active-Nickel-Core Cathodes

The chemical complications of an oxygen attack on an oxide cathode increase progressively with cores of platinum, pure

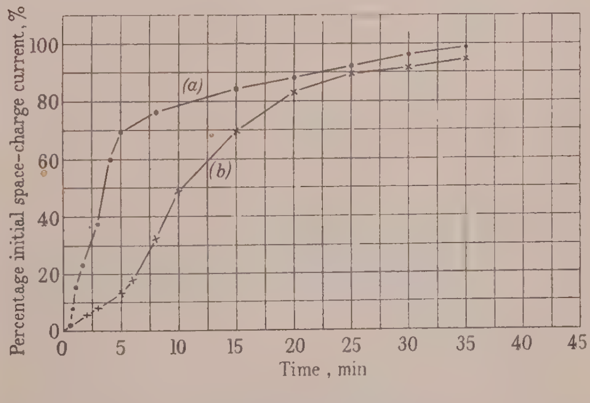


Fig. 26.—Recovery of active nickel- and platinum-core cathodes from a mild attack of oxygen.
(*a*) Platinum.
(*b*) Active nickel.

The oxygen attack in the above experiment is regarded as mild and well within the capacity of the gross reducing power of the active core.

(13.2) Comparison of Behaviour under a Heavy Attack

To demonstrate the relative response to a heavy oxygen attack the two type 6D15 diodes were subjected for 16 hours to a

* Kodak No. 3 density strip. Density increment, 0·15.

pressure of 1×10^{-4}mm Hg at 1 020°K. After pumping away the excess gas, the valves were sealed from the pump, the getters were flashed, and the cores were left to reactivate at 1 020°K under the current-carrying condition. Recovery was rapid and complete in the platinum case but notably absent on the active nickel core. Extensive treatment at 1 250°K also failed to improve the nickel core, and it seems clear that the nickel-oxide interface is more than chemically equivalent to the gross core reducing power. If the core is finally taken up to 1 400–1 500°K it tends to discolour—probably a mild example of a nickelite reaction.

(13.3) Influence of Cathode Current on the Reactivation of an Active-Core Cathode

A group of three active-core type 6D15 diodes was activated on the pump and subjected to a mild oxygen attack at 1020° K under a pressure of $5 \cdot 0 \times 10^{-5}$mm Hg for 4 min. The gas was then pumped away and the valves were sealed from the manifold. Recovery was effected at 1 020°K with one control sample under current load and the other two with zero load. From time to time the two test valves were explored for reactivation state by a 1 sec pulse method at the same potential as that employed in the control case. The results are set out in Fig. 27, and they show

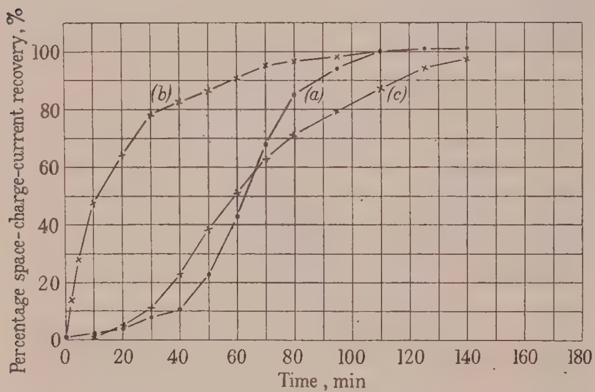


Fig. 27.—Influence of cathode current on recovery of active-core cathodes.

(a) Continuous cathode current.
b) and (c) No cathode current.

that recovery is largely independent of current flow, and with the two test valves it may be regarded as a simple chemical reducing action. It will be recalled from Part 1 that in the platinum case no such recovery is possible, and quick reactivation is basically dependent on current flow.

(13.4) Nickelite Reaction with Active-Core Cathodes

A useful impression of the role played by the core activator in the recovery of active-core cathodes from oxygen attack can be gained by comparing the nickelite reactions on passive and active nickel cores. Two groups of type 6D15 diodes were processed, subjected to 16 hours at a pressure of 1×10^{-4}mm Hg at 1 020°K and sealed from the pumps after the firing of the getter. Both groups were then run at 1 250°K for two hours to enable the core activators to work on the nickel-oxide interface. The cathodes were finally taken up to 1 400–1 500°K, whereupon the passive group showed an intense purple nickelite reaction whilst the active group gave only a faint discoloration. It is imagined that the active cores effectively reduced the nickel-oxide layer.

(13.5) Vulnerability to Oxygen Attack

It has been shown that platinum is greatly superior to both nickel-core types in recovery from oxygen attack, and it might be assumed that the platinum-core valve is relatively careless of gas and easy to process. However, this is not the case, and Fig. 28

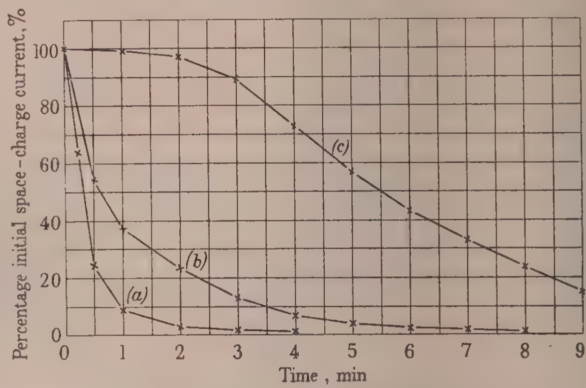


Fig. 28.—Response to common oxygen attack.

(a) Platinum core. No reducing power.
(b) Passive nickel core. Low reducing power.
(c) Active nickel core. High reducing power.

shows that platinum is much more vulnerable to a common oxygen attack than either passive or active nickel cores. The reason for this lies clearly in the absolute dependence of the platinum core on electrolysis, whereas the two nickel cores are able to regenerate barium to varying extents independent of the existence of cathode current. It follows that, in early life at least, the active nickel core can maintain a common level of emission at a higher residual gas pressure than can the platinum core. This insistence on a high quality of vacuum for efficient operation is a characteristic of platinum-core valves, and probably explains the difficulty that experimenters have encountered in preparing such valves with high emission levels.

(13.6) Comment

The passive nickel core was seen to differ from platinum in its prevalence to core oxidation and its general inability to clear itself of its nickel-oxide interface layer. The addition of core reducing power has enabled the active nickel core to cope much more successfully with this layer and to bring its buoyancy under oxygen attack more into line with that of platinum. However, there is a definite limit to the extent of oxygen attack that the active core can withstand, whereas there appears to be no reasonable limit with platinum. This limit is probably reached when the nickel-core layer is chemically equivalent to the core reducing power.

In a well-processed gas-free valve some core oxidation may occur during manufacture, but this is readily cleared, leaving the bulk of core reducing power available for reduction of the matrix itself. The resulting action at 1 020°K is a powerful flow of metallic barium, which maintains emission at an abnormally high level. Between 3 000–6 000 hours this action slows down, and the emission begins to fall steadily until after 15 000 hours it reaches the same adequate steady working level as that of a similar platinum-core cathode operating under the same current-carrying condition.

The virtues of this heavy flow of barium in early life are difficult to assess. The abnormally high level of emission during the first 5 000 hours leads to no material improvement in performance of the common receiving valve and represents a prodigal expenditure of barium. In a badly processed valve

suffering from high residual gas level the superfluity may have merit in keeping the cathode in operation for a few thousand hours; under gas-free conditions, however, the flow can be disadvantageous since it covers the insulating surfaces of the valve with a conducting film of barium and gives rise to electrical leakage, growth of stray capacitances and to intermittent noise.

The reliance to be placed on core reducing power as an aid to valve life over 100 000 hours is problematical. The power suffers decay, and at 1 020° K it seems probable that the half-life is under 10 000 hours with activity approaching a theoretical exhaustion at 50 000 hours. If the cathode temperature drops to 900° K, the rate of deployment of the activator is substantially reduced, and under such circumstances might be a valuable aid in the 100 000 to 200 000-hour range.

(14) CONCLUSIONS

The long-term emission behaviour of an oxide cathode has been examined in a broad rather than detailed fashion, and a generally consistent pattern of behaviour has been observed. The basic model now accepted by the author is a matrix running under gas-free conditions, losing barium by evaporation and regaining it by electrolysis. In such a system both the working level of emission and the rate of barium evaporation are functions of cathode current, and in the equilibrium state atoms of barium and negative ions of oxygen leave the cathode separately but in equal numbers. Chemical activation plays no essential role in the model, and its provision is regarded only as a technical aid to the cathode in its combat with residual gas in early life.

Perhaps the most interesting impression gained from the work has been the relative simplicity of oxide-cathode behaviour in the absence of a confusing core chemistry. It has, for example, been shown that the barium–strontium-oxide matrix is largely immune from permanent damage by molecular-oxygen attack, and after restitution of the vacuum condition it shows a natural inclination to reassume its useful active state under the impulse of current flow.

This is essentially an interim report, and the only justification for the emission-life model at this stage is that it seems in harmony with known facts.

(15) ACKNOWLEDGMENT

Acknowledgment is made to the Engineer-in-Chief of the Post Office for permission to make use of the information contained in the paper. The author also wishes to thank Mr. H. Batey for skilled assistance throughout the work.

(16) REFERENCES

(1) EAGLESFIELD, C. C.: "Life of Valves with Oxide-Coated Cathodes," *Electrical Communication*, 1951, **28**, p. 95.
(2) METSON, G. H., WAGENER, S., HOLMES, M. F., and CHILD, M. R.: "Life of Oxide-Cathodes in Modern Receiving Valves," *Proceedings I.E.E.*, Paper No. 1137 R, June, 1951 (**99**, Part III, p. 69).
(3) METSON, G. H., and HOLMES, M. F.: "Deterioration of Valve Performance due to Growth of Interface Resistance," *Post Office Electrical Engineers' Journal*, 1953–54, **46**, p. 193.
(4) METSON, G. H.: "Activation of High Vacuum Oxide-Cathode Valves," *Vacuum*, 1951, **1**, p. 283.
(5) FORMAN, R., and ROUSE, G. F.: "Oxide Cathode Base Metal Studies," *Journal of Research of the National Bureau of Standards*, 1951, **46**, p. 30.
(6) BECKER, J. A.: "Phenomena in Oxide-Coated Filaments," *Physical Review*, 1929, **34**, p. 1323.
(7) JOHNSON, C. B.: Unpublished work.
(8) METSON, G. H.: "Vacuum Factor of the Oxide-Cathode Valve," *British Journal of Applied Physics*, 1950, **1**, p. 73.

DISCUSSION BEFORE THE RADIO SECTION, 20TH APRIL, 1955

Dr. D. A. Wright: The paper presents some new results of considerable importance in a very clear and interesting manner, and it is very valuable, in particular, to have the behaviour of coatings of platinum established so definitely. I would like, however, to make a few critical remarks on some matters of detail.

First, my own experience comparing pure and active nickel does not entirely agree with that given in References 4 and 5. We have found that if we give the minimum processing treatment which is sufficient to produce optimum emission from coatings on active nickel, the same processing treatment, applied to coatings on pure nickel, does not in general produce such a high emission. Moreover, at all earlier stages of processing the emission from the pure nickel is lower. It should be noted that this is not a statement that pure nickel gives lower emission than active nickel; further activation will improve the emission from coatings on pure nickel, and it may ultimately become comparable with that from coatings on active nickel.

We have convinced ourselves, too, that with active nickel the coating in the best state of emission that we can produce has an optimum surface covering of barium, and has distributed in the coating also what seems best described as an optimum barium concentration. With the pure nickel cathode, the concentration of barium, both in and on the coating, is less than with the active nickel after a treatment identical with that which suffices for full activation of the latter.

These remarks are not very helpful, of course, unless one can suggest some reasons for the difference in results, and at present I am not able to do this. It would have been helpful, in discussing such differences, if in Section 8 the temperatures had been given at the different voltages in the processing treatment, i.e. 7, 10, 11 and 12 volts.

However, in view of the difference in results, I feel doubtful about the interpretation which the author gives of Fig. 19, i.e. in terms of conditions in which too much barium is present in the coating.

I agree completely that, by distillation of barium from an external source on to the coating, one can lower the emission, by forming too much bulk on the surface. But we have never found evidence that this greater-than-optimum concentration can be produced by thermal treatment of the coating itself or by drawing current from it. Of course, this does not prove that it cannot happen, but we have never found it to happen.

Secondly, I am uneasy about the statement in Section 2.3 that emission is directly proportional to conductivity. Incidentally, this statement would contradict the suggestions in Section 10.4 about what could happen if too much barium were present. Under many conditions this statement is true, but it need not always be true. The requirement is that the situation at the surface remains constant while the internal barium concentration is altering. If the situation at the surface is constant, both the conductivity and the emission are determined by the position of the Fermi level. But if the surface adsorbed layer alters as well as the position of the Fermi level, there is one factor affecting conductivity while two factors are affecting emission, so that there is not necessarily proportionality between them.

This fact will not, of course, alter any of the author's conclusions in the earlier part of the paper. However, it does affect

the argument in Section 10.4, where he discusses the magnitude of the barium loss, comparing platinum with pure nickel. For example, at some stage the emission may be falling during life, entirely owing to a change in the surface covering of barium, and not affecting the bulk concentration at all.

Under such circumstances a change in the surface concentration, linear with time, might produce a change in the work function, also linear with time, and hence a change in the emission, which would then be exponential with time. There could be quite a large change in emission with only a small change in the overall barium concentration. If this occurred, it would invalidate some of the remarks in Section 10.4.

The fact that the life curves are exponential, as shown in Fig. 5, does not, of course, prove my point. Other mechanisms might produce an exponential curve. However, it is clear that the possibility I have outlined should not be ignored.

Finally, I do not know why the author uses the expression "abnormally high emission level" in Section 13.6. The emission from the active nickel is not any higher at its best than the emission from platinum or pure nickel, according to his own figures. The highest emission which he quotes—and which incidentally agrees very well with the emission that we observe from well-activated coatings under good vacuum conditions—is typical of a coating whose surface is covered with barium, and in which the internal concentration of barium is the optimum at the temperature in question, in vacuum. I do not think that, in any sense, this emission can be described as abnormally high.

Mr. F. G. Haegele: The paper raises the fundamental idea that platinum simplifies the situation by avoiding either an oxidizing or a reducing interface, and it appears rather unfortunate in this particular case that we have found a new property for cores—that of core absorption and core dispensation. Can some of the difference which we have ascribed to this simple profit-and-loss account be ascribed to this new property rather than to the elimination of one of the old properties?

In Fig. 19 the two curves for passive and active nickel show manifestly the same peaks, and we can reasonably assume that these peaks are typical of an impurity in the cathodes and that they shift back in time, i.e. to a smaller number of hours, as the impurity level gets smaller. The author has posed the simple experiment of boiling out some of the activator. In view of the ratio of the impurities in passive and active nickel, which is of the order of 1 : 50, can we regard this simple experiment as significant, bearing in mind that as we remove the magnesium we can expect that the peak will travel to the left of the curve and the slope on the right-hand side of the curve will gradually diminish? The resulting change will be so small that it is doubtful whether, unless we have removed a very large proportion of magnesium, we should see any difference.

In examining these curves I feel that the left-hand side is the most noteworthy, and it would be interesting if the nickel manufacturers would provide extremely pure material, to see whether this upward slope could be made to turn itself inside out and become the same as curve (*a*).

I should also like to comment on the novel explanation of the difference between curve (*a*) and the curves for nickel. A small change in processing, which might diminish the quantity of barium, would provide us with more evidence. Has the author made this experiment? Alternatively, of course, there is the possibility of making a Richardson plot.

In Section 12.3.2, does the spectrographic analysis for magnesium refer to a test taken on the surface of the cathode or on the bulk of the metal? In the latter case it would depend very much on the thickness of the cathode sleeve. Under item (*a*), X-ray diffraction might prove the existence of the oxide form.

It may, of course, be no more than a coincidence that the steep fall in Figs. 24 and 25 corresponds in time to the rise in emission in Fig. 19 for the same active nickel. The author has stated that the results which are obtained here are quite slight in character, and I look forward to rather more data to cover this interesting period between 0 and 1000 hours, which may provide the pointer to further and more close interpretation of what I regard as the most interesting part of the zero-load characteristic curve.

Dr. R. O. Jenkins: The author is a little unjust to previous workers in stating that the concept of electrolysis has never been proved, and that the proof would be the evolution of free barium rather than oxygen. In 1932 Berdennikova in the Soviet Union actually carried out the experiment. I believe that he correlated the total free barium evolved by a cathode during running with the emission (in coulombs) that had passed through the coating. His method was to have the cathode and anode running in a valve connected to a pump. When the valve was evacuated and outgassed it was operated with anode current flowing. Finally, the free barium was estimated by letting in water vapour and the resultant free hydrogen, produced by reaction with the free barium, could then be measured after the water vapour had been frozen out.

We repeated some of these results six or seven years ago and extended them, because originally only the total amount of barium produced was examined. Some of it was left in the matrix of the cathode and some was, of course, evaporated over the anode and possibly also on to the glass.

We wanted particularly to find separately the amount retained in the matrix and also the amount which evaporated. If we wanted to measure the amount on the anode it was dropped off into a side tube and the valve was removed from the pump, and similarly for the cathode. That also enabled us, although we were working on a pump system, to have a getter in the valve and approximate to normal valve conditions. The vacua were not, however, quite as good as in a commercial valve and definitely not as good as in the author's carefully processed valves.

One interesting point emerged when we re-examined the results. We had tested spectroscopically pure nickel and platinum cathodes. They were given identical treatment during the activation and the same emission (in coulombs) during ageing for about an hour before estimating the barium evaporation. The evaporation from the pure nickel corresponded to just over a microgramme of pure barium, whereas from platinum it was only about $0 \cdot 6 \mu g$. Neither of these cathodes could have produced any free barium owing to chemical means. Therefore it must have been produced by electrolysis, and one would have expected the same quantity to be produced. It appears that the odd $0 \cdot 4 \mu g$ had gone into the platinum, and this to some extent confirms the author's results.

It would be interesting to take some of the cathodes which the author had been using and strip off the coating. Instead of measuring emission with time and determining the Q-value, which is an integral of time emission, could not a tungsten filament be mounted adjacent to the platinum and the amount of barium being emitted be measured by the time taken to deposit a monomolecular layer on to the tungsten?

A fairly close estimate could then be made of the total barium in the platinum. According to our measurements it would, after initial activation, be rather less than a microgramme. Assuming the figures given for the Q-value held, it would appear that about twenty times that amount might come from a valve that had been running for a long period with space current.

Monsieur A. Bonan (*France*): The author states that for good valve life we have to consider emission and interface resistance. He seems to connect the two phenomena with the absence of silicon, which is why he tried to use platinum.

In Reference 3 the author refers to the use, in receiving-valve manufacture, of a nickel containing about 0·05% magnesium and 0·08% silicon. During the processing the silicon reacts with the oxide matrix to form an interface layer of barium orthosilicate. At the beginning of the life we do not see anything, because this layer is activated by an excess of free barium atoms. The progressive deactivation of this layer owing the destruction *in situ* of the free barium by an invading oxidizing gas leads to the rise of interface resistance.

In the present paper the author states that, if active nickel containing silicon and magnesium is used, we have, at the start, reasonable emission. If a small amount of oxygen is present the emission will fall. If we remove the gas the reducing agents come into play, and we have recovery of emission, because of the recovery of free barium atoms. But if free barium atoms appear in the layer they should also activate the orthosilicate and remove the interface resistance. However, if we make the test, we find, after a time, that interface resistance occurs.

Emission does not seem to bear a direct relation to interface resistance. We have had valves out of action owing to high interface resistance, which had higher emission than at the beginning of their life, before interface resistance was present.

If the emission of a tube is relatively steady, the same cannot be said of the interface resistance.

We found valves, out of action, which had an interface resistance of about 100 ohms, and we thought that it might be possible to compare them from time to time with other valves of the same kind. We put them on one side, and after three months we tested them again. The emission had not changed, but we found no interface resistance. It seems that Bounds and Briggs* also found such phenomena.

According to the author's reasoning, in the first case interface resistance should not appear, but it appears. In the second case it should remain constant, but it disappears. Moreover, in the two cases, the emission seems to be unconnected with interface resistance.

THE AUTHOR'S REPLY TO THE ABOVE DISCUSSION

Dr. G. H. Metson (*in reply*): I thank the speakers for their comments, which have been stimulating and thoughtful. Dr. Wright's views on the emission levels from pure and active nickels are, to my mind, identical with my own. Under good laboratory conditions the levels are about equal; under poor processing conditions the active nickel is always in better shape to weather the gas storm. I have sympathy with Dr. Wright's doubts on the explanation given to account for the rising emission characteristic in Fig. 19. The idea of a "greater than optimum" barium concentration in early life is perhaps unconventional, but it is at least a rational hypothesis and I note that Dr. Wright offers nothing to supplant it. The paper makes considerable use of the supposed proportionality of surface emission and matrix conductivity, in that it assumes that total emission is a measure of excess barium content of the matrix. I agree entirely with Dr. Wright that this proportionality can be drastically disturbed by surface gas attack, but much care was taken in the work to see that the cathode was in concentration equilibrium at any time of emission measurement. Much of the work was carried out with the cathode under zero load in a sensibly gas-free atmosphere, and it is difficult to imagine a concentration disequilibrium occurring under such a condition.

Mr. Haegele raises the point that platinum-core absorption and dispensation might complicate oxide cathodes rather than simplify them. I disagree with this contention, since in the equilibrium (or long-term) state the two actions are self-balancing and can be ignored. I am strongly of the opinion that our ultimate understanding of the oxide cathode will come from platinum-core studies, where the cathode can be examined in an isolation uncomplicated by inessential chemical side issues. The spectrographic studies of Section 12.3.2 for magnesium were conducted on the bulk material.

The points made by Dr. Jenkins on the subject of electrolysis are pertinent, but I am surprised that he failed to raise the question of disposal of the electrolytic oxygen or the influence of electrolysis on cathode life. We are a little at variance in regard to our definition of "proof," but I think that this is merely a question of attitude based on relative circumstance. In industry the active nickel core makes the existence of the electrolytic mechanism a matter of academic importance only, but the user of a platinum-core valve is wholly dependent on the phenomenon. With the bulk of British submarine repeaters working with platinum-core pentodes it is perhaps natural that at the Post Office we should remain excessively critical of the phenomenon.

I agree with almost all that Monsieur Bonan has said. The essentials of long mutual-conductance life are the maintenance of adequate emission and the absence of interface growth. The latter can now be avoided, and I think there is little profit in studying it further. Future effort should concentrate on the basic problems of maintenance of emission.

* *Le Vide*, May, 1954, p. 5.

621.385.032.216 : 621.315.612.8 : 621.3.015.5

Paper No. 1857 R
Sept. 1955

SOME EXPERIMENTS ON THE BREAKDOWN OF HEATER-CATHODE INSULATION IN OXIDE-CATHODE RECEIVING VALVES

By G. H. METSON, M.C., Ph.D., M.Sc., B.Sc.(Eng.), Associate Member, E. F. RICKARD, B.Sc., and F. M. HEWLETT, B.Sc.

(*The paper was first received* 20*th November*, 1954, *and in revised form* 8*th March*, 1955.)

SUMMARY

Sintered alumina used for the insulation of valve cathode heaters is preferentially liable to sudden failure in service when the heaters are at a positive potential with respect to the cathode. The phenomenon has been studied in conventional valves run in varying conditions and also, in more detail, on simple two-electrode systems using recrystallized alumina as the insulant.

The sudden failure is shown to be the last of three stages, (*a*) a time-consuming degradation of the insulant with rate exponentially related to operating temperature, (*b*) a period of intermittent sparking and arcing, and (*c*) a local weld between heater and cathode which appears as a more or less permanent circuit fault.

The initial stage is accompanied by a transfer of tungsten from the heater across the alumina, and it is this process which is strongly enhanced when the heater is the positive electrode. Production of positive tungsten ions under electron bombardment from the cathode is postulated, and it is shown that tungsten transfer and also insulation failure are greatly retarded when the alumina contains a pore-free layer. This may be achieved in the laboratory by surface fusion or by the use of monocrystalline sapphire.

Reasons for the apparent absence of widespread heater-cathode insulation failures are suggested.

(1) INTRODUCTION

(1.1) General Nature of the Insulation Breakdown

Cathode heaters normally consist of a fine tungsten wire formed in some suitable shape and covered with a thin layer of sintered alumina which insulates the heater from the interior of the nickel cathode core. For reasons of circuit design the heater and cathode core may be operated at different potentials, and a failure of insulation under this condition may give rise to circuit derangement.

Experience over the past few years seems to indicate that the probability of breakdown is related to the direction of the applied potential, and the dangerous direction is that of heater positive to cathode. When breakdown occurs it appears to be quite sudden, with a fall of insulation from a thousand megohms to a fraction of a megohm occurring in a time of less than a minute.

The paper describes the breakdown occurring in a range of common receiving valves, and records some preliminary experimental observations on the course and nature of the phenomenon.

(1.2) Scope of the Investigation

The investigation has been divided into two parts—a phenomenological inquiry into the breakdown of common receiving valves, and an experimental approach to the nature of the phenomenon through a simplified heater-cathode system.

(2) BATCH TESTS OF COMMON RECEIVING VALVES

(2.1) Testing Technique

The valve type selected for examination was a high-slope indirectly-heated pentode with a 2-watt cathode system. This particular type was adopted as it is in common use, it is made in the United Kingdom by all of the large valve manufacturers and it can be readily assembled in the laboratories at the Post Office Research Station. Test batches of a common type were therefore available from a wide range of manufacturing sources, and it was furthermore possible to vary the heater-insulant system over an experimental range in the batches assembled at the Research Station.

The general method of batch testing was to mount the test group on a substantial rack immune from mechanical shock and to run the group at some selected voltage with the heaters maintained at a positive potential of 36 volts over the cathode cores (this particular potential was selected as of interest to the Post Office in the operation of submarine repeaters). Breakdown of heater-cathode insulation was detected and recorded by automatic repetition searching over the valve bank, each position being tested at intervals not exceeding thirty minutes.

(2.2) Treatment of Data

(2.2.1) Distribution of Lifetimes within a Batch.

A study of the time to first failure of valves, within a batch tested in particular conditions, showed them to fall reasonably well into an exponential distribution with, consequently, a probability of failure constant with time. Results for one batch of 44 valves, shown graphically in Fig. 1, illustrate the extent of agreement between actual and theoretical distribution of survival for such an exponential distribution.

For an infinite population, the time-constant of the exponential may be determined at any stage as the sum of the lives of the failures and of the testing times of the survivors, divided by the number of failures. The corresponding calculation of average life of valves in a sample batch at any stage gives the time-constant for the population life-distribution curve within confidence limits which may be derived from the number of failures and the Poisson probability distribution.

(2.2.2) Batch Average Life.

The time-constant of the assumed exponential distribution curve, calculable within known confidence limits, may be taken as a measure of batch average life without the necessity for testing to failure the entire batch.

Calculated for 95% confidence limits, it has been used as the criterion for assessing batches of valves run in different conditions.

(2.3) Influence of Various Factors on Batch Average Life

(2.3.1) Heater Temperature.

Batches of valves were tested with different heater voltages. Heater temperatures were estimated from characteristics of heater current and voltage, using Shardlow's curves of resistance and temperature. The plot shown in Fig. 2 indicates a reasonably linear relationship between the logarithm of batch average life and the reciprocal of heater temperature. There was no indication that the nature of the breakdown had been altered at the

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
The authors are at the Post Office Research Station.

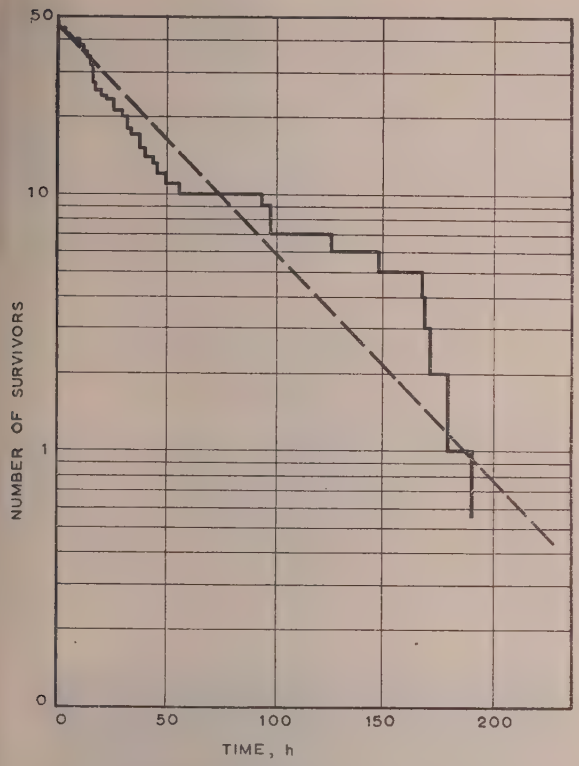


Fig. 1.—Distribution of survival within a batch of valves.
—— Curve showing actual survivors.
---- Theoretical curve.

higher temperatures, and many subsequent tests were accelerated by running the heaters with a 10-volt supply, giving a heater temperature of near 1 700° K.

(2.3.2) **Magnitude and Direction of Applied Direct Potential.**

All systematic testing has been carried out at the constant potential difference of 36 volts, but isolated tests at other potentials have left the impression that an increase in applied potential tends to shorten the average life.

Table 1

	No. of failures	Average life	Batch average life (95% confidence limits)
		hours	hours
Heater positive	46	50	38–68
Heater negative	10	6 340	3 450–13 210

Evidence of a directional effect is given in Table 1 relating to two batches of valves tested at 1 700° K with a breakdown potential of 36 volts.

(2.3.3) **The Cathode Core and the Heater Metal.**

No evidence was obtained to show that life was affected by the use of platinum, nickel or molybdenum as the cathode core metal.

Also, no evidence was obtained to show that life was affected by the use of several varieties of silicated tungsten and molybdenum-tungsten as heater material.

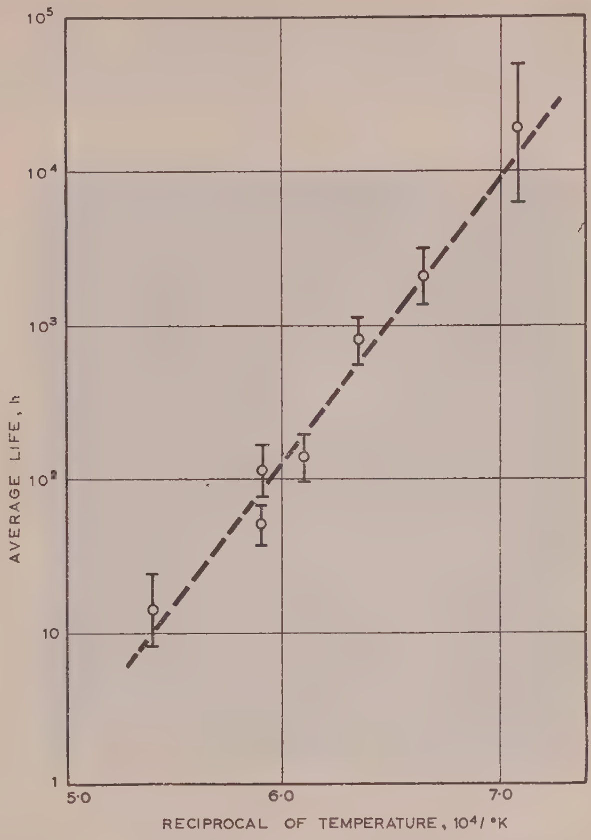


Fig. 2.—Batch average life as a function of temperature.
I 95% confidence limits for average life.

(2.3.4) **Heater Coating.**

Various commercial brands of alumina powder from English, American and German sources were made into spraying pastes and electrophoretic coating suspensions. Pastes used in the commercial coating of heaters were also tested. Heaters coated with these were sintered at different temperatures and mounted in valves. No marked variation in batch life was found, but there was an indication that some powders give slightly longer life when sintered in the range 1 650°–1 700° C in preference to temperatures above 1 700° C.

(2.3.5) **Heater Construction.**

One batch of valves had an unusually long life when tested with occasional inspection, but when the survivors were opened and examined, it was found that breakdown had occurred undetected. These heaters were of the "faggot" type, all others being folded spirals, and the apparent improvement in life was attributed to the facility of free movement of this type of heater. A repetition of the tests, under continuous inspection, gave a much shorter life.

(2.4) **Breakdown Sites**

(2.4.1) **Visual Appearance.**

After breakdown, heater coatings showed small stained patches the appearance of which suggested that they may have been due to arcing between heater and cathode. There was often a central dark spot with a metallic glint which lay at the bottom of a minute crater, and the inner surface of the cathode sleeve was

roughened and stained in patches corresponding to the positions of the stains in the alumina. No connection could be established between visible physical imperfections in the original coating and the position of the stains, which were taken to be the breakdown sites. Fig. 3 shows part of a conventional heater after breakdown.

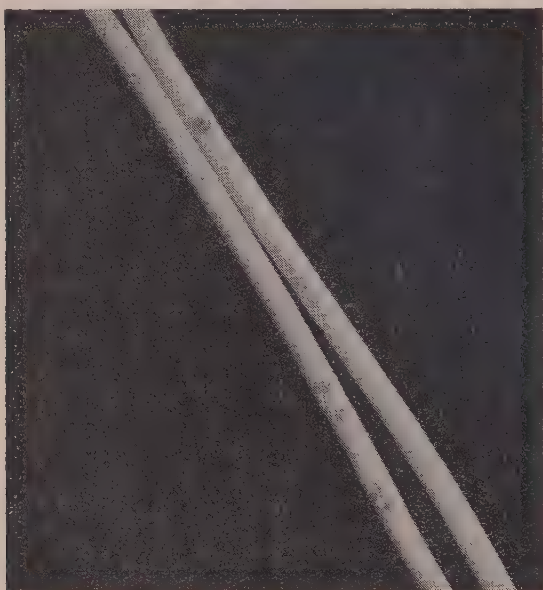


Fig. 3.—Conventional heater showing breakdown sites.

(2.4.2) **Stages of Growth.**

A survey of many scores of stains suggested that they had developed in a consistent sequence as follows:

Stage (*a*).—A small grey patch of indefinite shape.

Stage (*b*).—Enlargement and assumption of a well-defined, roughly circular shape of even grey.

Stage (*c*).—An intense stain at a central spot surrounded by an area of pale grey, bounded by a thin line of darker grey.

In a batch of valves run for constant time, the numbers of stage (*c*) stains on the heaters were roughly in the inverse ratio of the recorded times to first breakdown.

(2.4.3) **Chemical Tests.**

Chemical examination of material removed from stained portions of the heater showed tungsten present in all stains tested, and cathode core metal in some. There was always some staining material which could not be identified, being embedded in fused alumina grains, and this could have been heater or cathode metal or a compound of aluminium.

(2.5) Summary and Comments on Batch Tests

The results indicate that there is, in all the conventional heater-cathode assemblies tested, the liability of sudden breakdown of insulation. There is a marked directional effect, the liability being great when the heater is at positive potential with respect to cathode. This liability is an exponential function of operating temperature. The distribution of lives of valves within a batch approximates to an exponential distribution, which gives a constant probability of failure with time.

The nature of these batch test results can be correlated with the appearance and frequency of the breakdown sites if it is assumed that an initial time-consuming process leads to degradation of the insulant with consequent production of transient sparks or arcs which would not normally be detected by the intermittent testing mechanism. Movement of the heater would often prevent this appearing as a permanent fault, but the worst stains are seen when local adhesion has occurred. Such an event would cause the fault to remain until severe movement, e.g. the mechanical vibration of the whole valve, disturbed the contact. It was, in fact, found useless either to remove the valve or to reduce the temperature of the heater in attempts to measure the cold resistance of the path, since the local short-circuit generally disappeared.

(3) TESTS ON SIMPLIFIED SYSTEMS

(3.1) Technique

(3.1.1) **Description of Assembly.**

Commercial recrystallized alumina tubes, of bore 0·75 mm and wall thickness 0·43 mm, were cut to length so as to take a conventional 2-watt tungsten spiral heater, straight and uncoated. Around the exterior of the tube was wound in close contact a spiral of tungsten or platinum wire to serve as a second electrode. This assembly was mounted on a standard valve pinch, sealed in and pumped to high vacuum. Fig. 4 shows a typical assembly.

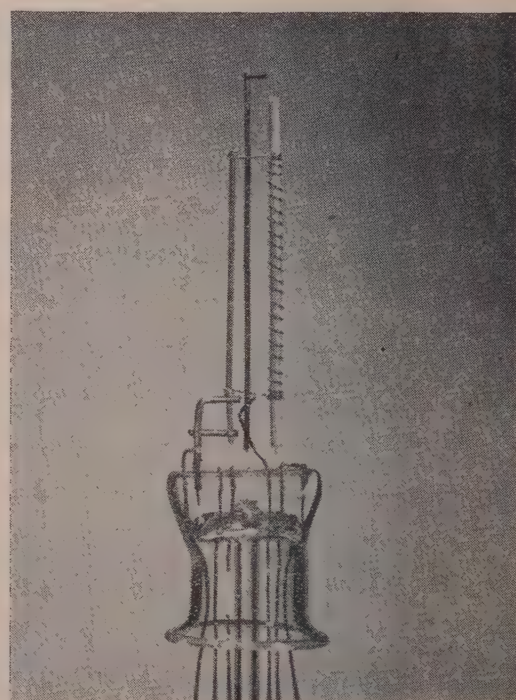


Fig. 4.—Simple assembly using recrystallized alumina tube.

(3.1.2) **Method of Test.**

Heater temperatures were adjusted by varying the voltage and were in general much higher than in the tests on normal valves—usually about 2 300° K. The surface temperature of the alumina tube was approximately 1 700° K measured by an optical pyrometer. The applied potential was 500 volts from a d.c. source, equivalent to a voltage gradient across the alumina of 10 kV/cm as compared with 5 kV/cm in the tests on normal

valves. Leakage currents were measured by a suitable meter protected by a 0·5-megohm series resistor.

(3.2) Test Results

(3.2.1) Leakage-Current Characteristics.

The variation of current with time was studied, and a typical graph is shown in Fig. 5. An initial small rise was followed by a

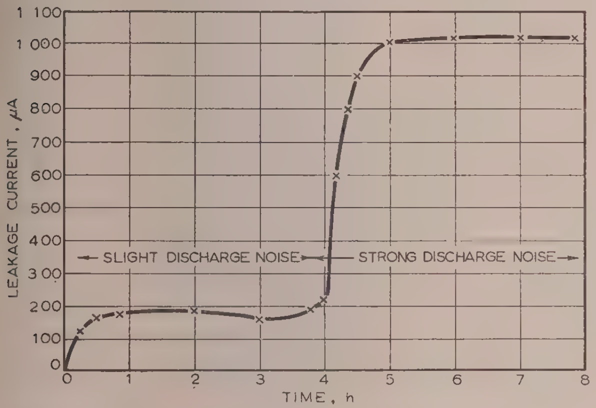


Fig. 5.—Current/time characteristic under breakdown-inducing conditions.

period of about 10 hours in which the current was stable. The current then rose rapidly, with violent fluctuation, to the value set by the limiting resistor, and this level was maintained for an ill-defined period terminated by heater burn-out. The curve of Fig. 5 has been smoothed to show the general trend in the rapidly rising period. During the initial stable period, it was possible to study the current/voltage relationship, and Fig. 6 shows this characteristic.

(3.2.2) Discharge Noise.

When a pair of high-impedance headphones was inserted in the d.c. circuit, it was noted that there was noise in the form of "clicks" superimposed on steady "frying." If the current was increased during the stable period by an increase in applied voltage, there was a rise in general noise level. During the unstable period, the rising current again produced a rise in background noise level and also resulted in increased frequency and loudness of "clicks."

(3.2.3) Breakdown Stains.

During breakdown, stains appeared on the outer surface of the alumina and developed roughly according to the scheme postulated in Section 2.4.2. These gradually grew together, producing a broad spiral band along the line of the external cathode wire. After breakdown, the alumina tube was sectioned and ground on Carborundum. Inspection of the stain shapes at different stages of grinding suggested that the stains had grown radially through the alumina from external cathode to internal anode (i.e. heater). Fig. 7 shows a typical stained tube and the same in cross-section.

(3.2.4) Effect of Reversed Potential.

Two tubes, both carrying tungsten heaters and tungsten spiral outer electrodes were run at the same temperature and applied potential. In the system where the inner, hot wire was positive, catastrophic breakdown occurred in about 10 hours. Where the outer, cooler wire was made positive no breakdown occurred in

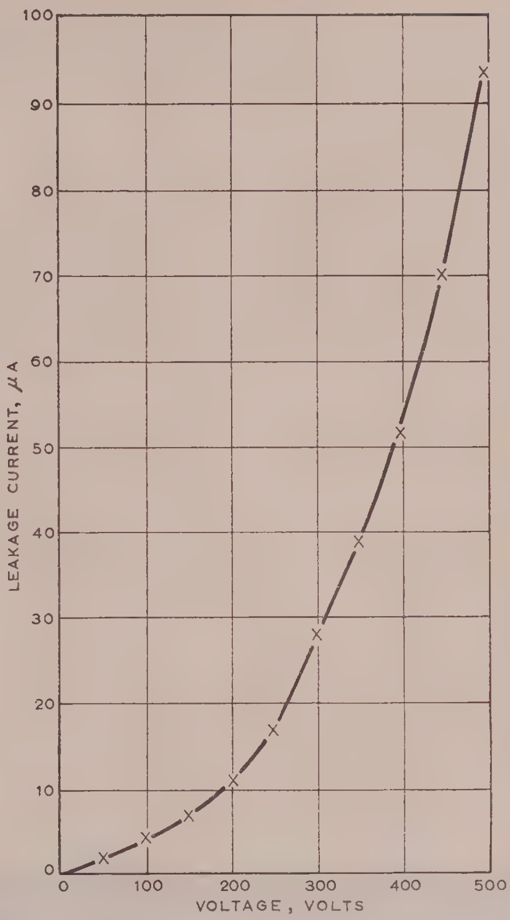


Fig. 6.—Current/voltage characteristic before breakdown.

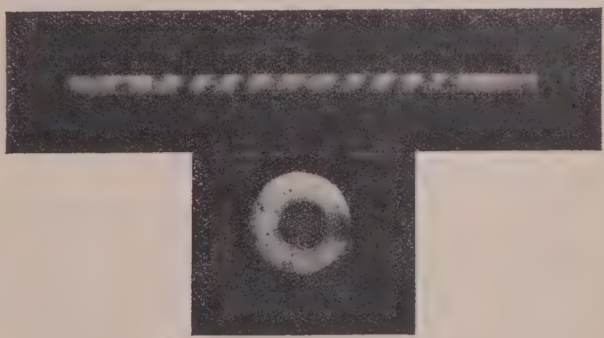


Fig. 7.—Tube and cross-section showing breakdown stains.

100 hours, and though there was some general discoloration of the alumina in the tube, there were no typical breakdown stains.

(3.2.5) Effect of Lowered Cathode Temperature.

Some assemblies were made in which the spiral cathode in close contact with the alumina was replaced by a hemicylinder of platinum foil separated from the alumina by a gap of about 1 mm. These were run with heater temperatures of 2 350° K.

surface temperature of the alumina was approximately 1 800° K, but the temperature of the foil probably did not exceed 600° K. The potential was 500 volts, with heaters positive. After 300 hours there was no detectable leakage current and no stain formation.

(3.2.6) **Tungsten Transfer Phenomena.**

Fairly dense films of tungsten had often been found on the inner alumina surface of tubes and of normal heater coatings. In the tests described in Section 3.2.5 the bores were unusually clean, being practically free of tungsten despite the high temperature. A new test was made in which pairs of tungsten heaters were run at the same temperatures in the bores of small twin-bore alumina tubes with 500 volts applied potential between the heaters. One set was operated at 1 600° K and the other at 2 300° K. In each case the bore surrounding the positive heater was contaminated with considerably more tungsten than was present in the bore of the negative side. It is also of interest that the "bores" in the sapphire plate assemblies described in Section 3.3 were nearly free from tungsten film after 1 500 hours at 2 000° K. In those tests the cathode was hot, about 1 400° K, but passage of current was prevented by the nature of the insulant. These results suggest that bombardment of a heater by electrons from the cathode is probably a potent factor yielding enhanced tungsten transfer.

(3.2.7) **Effect of Fusing the Alumina.**

Some tubes were heated in an oxyacetylene flame until the outer surface was glazed, presumably giving a pore-free layer. One of these and an unglazed tube were tested under the usual conditions. The untreated tube suffered catastrophic breakdown in 8 hours, whereas the glazed tube maintained a steady insulation resistance of 5 megohms for over 300 hours. The stains in the unglazed tube were manifold and had in some places completely penetrated the alumina. The glazed tube showed only a few stains of negligible penetration. Fig. 8 shows this tube and it may be compared with Fig. 7.

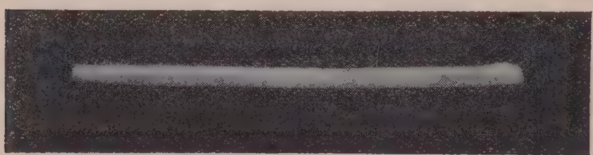


Fig. 8.—Surface-glazed tube showing slight staining.

(3.2.8) **Chemical Aspects.**

Tungsten emanating from the heater could always be detected in the stained portions of the tube, and in some instances it was also detected on a platinum outer electrode. When a platinum cathode was used, this metal also could be detected in the more intense stains. As with the ordinary heater coatings, there was present staining material within alumina grains which could not be identified.

(3.3) Use of Synthetic Sapphire

(3.3.1) **Description of Assembly.**

As an extension of the tests with glazed tubes, tests were made using monocrystalline alumina. Small plates of synthetic sapphire were grooved so that, when mounted in pairs, there was room for a conventional two-leg heater to lie in the grooves. The plates were bound tightly together with a spiral of platinum wire which served as cathode. The minimum wall thickness was 0·25 mm. The assemblies were run with the heater positive.

(3.3.2) **Test Results with Sapphire Plates.**

At first, the specimens were operated at heater temperature of 1 900° K and 500 volts from a d.c. supply. For 600 hours the insulation resistance remained greater than 10 000 megohms. The potential was then raised to 1 000 volts, equivalent to a maximum gradient in the alumina of 40 kV/cm, and the heater temperature was raised to 2 000° K. This was maintained for a further 1 500 hours during which no staining could be observed, and the insulation resistance remained greater than 5 000 megohms. The temperature was raised to 2 350° K, and after another 500 hours the heaters burnt out without any increase in leakage current.

(3.4) Comments on Tests with Simple Systems

These tests have given an insight into the nature of the processes occurring during the induction period. The chemical evidence suggests that transfer of tungsten into and across the alumina may be the rate-determining process during the induction period when gradual degradation of the insulant is thought to occur. The results with cool cathodes indicate that tungsten transfer itself is dependent on electron flow. The results with partly fused and monocrystalline alumina show that porosity is a necessary link between electron emission from the cathode and tungsten transfer into and across the alumina. They also show that ordinary electrolytic processes cannot play a major part in catastrophic breakdown.

(4) DISCUSSION OF RESULTS

The main results of these experiments may be summarized as follows:

(*a*) The breakdown, recognized in a working valve as a sudden failure of heater insulation, is the last stage in a process, the rate of which is an exponential function of temperature (Section 2.5).

(*b*) One feature of this process is the transfer of tungsten from the heater into and across the alumina (Section 3.2.8).

(*c*) Tungsten transfer is strongly dependent on the passage of current between cathode and heater (Section 3.2.6).

(*d*) The directional effect of applied potential observed in working valves is also observed in systems where both electrodes are of tungsten (Section 3.2.4).

(*e*) The leakage current and the transfer of tungsten are suppressed by the presence of a layer of non-porous alumina (Section 3.3.2).

An explanation of these facts can be given by assuming that positive tungsten ions are produced at the heater under electron bombardment. (These would move towards the cathode, and there is evidence that some tungsten actually reaches the cathode surface.) The rate of tungsten transfer is then dependent on rate of positive-ion production, and the latter depends on current magnitude and on the temperature of the positive electrode.

It is likely that some dissociation of the alumina itself, under electron bombardment, occurs simultaneously, and in this event oxygen ions will migrate towards the (anodic) heater and transfer of tungsten by migration as oxide is likely. The evidence does not permit a clear assessment of the importance of this stage.

The presence of tungsten in the alumina and of positive ions in the region of the cathode surface would both tend to an increase in local field-strength at certain points. Such an effect would be self-accelerating and would terminate in a breakdown arc, as is found to occur.

The breakdown can be avoided if the initial process, i.e. the bombardment of the heater with electrons, is prevented by interposition of a pore-free layer of alumina.

The mechanism outlined is undoubtedly simplified, but it probably includes the major factors responsible for heater insulation breakdown.

Having regard to the large number of series-operated a.c./d.c. receivers used on d.c. mains in this country with satisfactory results, it is curious that insulation breakdowns have not been more widely observed. In considering this point, the authors think it possible that the lack of widespread experience may be due to one or more of the following reasons.

(*a*) Electrical and mechanical life of the valve itself may in some cases be less than the insulation life of the heater.

(*b*) The breakdown "site" is unstable mechanically, and a low level of mechanical shock is usually sufficient to clear a fault. Examples have been found of valves with normal insulation but which have shown numerous breakdown sites when opened for examination. It is thought that such valves have suffered many heater insulation failures which have cleared themselves by mechanical shock derived, e.g., from normal handling or the thermal shock of heater voltage switching.

(*c*) A number of cases have been noted in which actual mechanical fracture of the tungsten heater wire has coincided physically with an insulation breakdown site. It seems reasonable to assume, therefore, that a proportion of heater open-circuit failures is due initially to the insulation breakdown phenomenon.

(5) ACKNOWLEDGMENT

Acknowledgment is made to the Engineer-in-Chief of the General Post Office for permission to make use of the information contained in the paper. The senior author also wishes to thank Mr. D. J. Ravenhill, who has been associated with the experimental work involved in the tests on simplified systems.

DISCUSSION ON "NOISE GENERATION IN CRYSTALS AND CERAMIC FORMS OF BARIUM TITANATE WHEN SUBJECTED TO ELECTRIC STRESS"*

Dr. R. Street (*communicated*): It has been shown by the author that the noise developed in ferro-electric crystals and ceramics can be ascribed to the occurrence of discontinuous changes in polarization, and the analogy to the Barkhausen effect in ferromagnetics has been pointed out. Attention has also been drawn to the time effects associated with noise generation by ferroelectrics. The similarity of the ferro-electric and magnetic cases may be even more striking than is claimed.

For ferromagnetic materials the change of the intensity of magnetization, consequent upon a stepwise change in the applied field, is time dependent.† The phenomenon has been termed magnetic viscosity or after-effect. It is generally accepted that the observed time dependence is due to the influence of thermal agitation on the domain processes. For example, under certain circumstances, thermal agitation may provide sufficient energy to activate the movement of a domain boundary from positions of metastable to stable equilibrium. The process is thus one of thermal activation of Barkhausen discontinuities. By making plausible assumptions it is possible to show that the time-dependent component of the intensity of magnetization should theoretically be of the form $I(t) = S \log t$.‡ The origin of time, t, is the instant at which the stepwise change in field is made. In simple cases, S is directly proportional to the absolute temperature and depends *inter alia* on the material of the specimen and on the point of measurement on the magnetization curve. The predicted form of $I(t)$ as a function of time has been observed experimentally for a number of materials, and, in particular, measurements of the quantity S have been made at different points on the magnetization curve. The maximum value of S for a given material occurs at, or near, the coercive force point of the magnetization curve. Thus the magnetic behaviour is similar to the ferro-electric behaviour illustrated in Fig. 11. Measurements on the temperature dependence of S, using specimens of Alnico and similar materials, demonstrated the linear relation predicted by theory. There are obvious difficulties in extending this to the ferro-electric case owing to the restriction of the available temperature range by the transition points and also by the increase in electric conductivity at the higher temperatures.

The majority of measurements on magnetic viscosity have been made by direct magnetometric measurement of the intensity of magnetization of the specimen. The slow response of the magnetometer makes it impossible to observe individual Barkhausen discontinuities while magnetic viscosity is proceeding. In the present connection it is interesting to note that the direct analogue of the ferro-electric method described by the author has been adopted in the investigation of magnetic viscosity of very-high-coercivity alloys.† In this form of apparatus the time differential of the intensity of magnetization is measured in terms of the e.m.f. induced in a search coil wound on a specimen placed between the poles of an electromagnet. For the majority of materials examined the e.m.f.'s induced by single Barkhausen discontinuities are not sufficiently large to be resolved as individual pulses; the observed e.m.f., V, is a smoothly decreasing function varying as $1/t$,

$$\text{for } V \propto \frac{d}{dt}[I(t)] \propto \frac{d}{dt}[\log t] \propto \frac{1}{t}$$

The failure to resolve the Barkhausen discontinuities is not surprising, since there is good evidence to show that for these specimens the volume of material involved in a Barkhausen discontinuity is about $10^{-17}\,\text{cm}^3$. However, with suitably heat-treated Alcomax specimens it is possible to observe directly the impulsive Barkhausen e.m.f.'s superimposed on a $1/t$ curve. The observed variation is essentially similar in form to that shown in Fig. 5.

It would be very interesting to know whether any results are available which could be used to demonstrate quantitatively the existence, or otherwise, of "ferro-electric viscosity." If the phenomenon does in fact exist there appears to be no major difficulty in applying the thermal activation theory, developed for magnetic viscosity, to the problem. This would be of value in that the results shown in Fig. 6 suggest that the processes involved may be of a particularly simple form. The movements of the domain walls are also directly visible as noted in Section 5.5.1. As is the case for observations on magnetic viscosity, it

* KIBBLEWHITE, A. C.: Paper No. 1752 M, January, 1955 (see **102 B**, p. 59).
† EWING, J. A.: *Proceedings of the Royal Society*, A, 1889, **46**, p. 229.
STREET, R., and WOOLLEY, J. C.: *Proceedings of the Physical Society*, 1949, **62A**, p. 562.
‡ STREET, R., and WOOLLEY, J. C.: *ibid.*, 1952, **65B**, p. 679.

† STREET, R., and WOOLLEY, J. C.: (To be published).

may be that the investigation of ferro-electric viscosity could lead to information on small-scale domain phenomena.

Dr. A. C. Kibblewhite (*in reply*): Unfortunately, few of the results obtained in the course of the noise investigation of the barium titanate group lend themselves conveniently to an examination along the lines desired. However, those which are available do appear to support Dr. Street's suggestion that there may be a close analogy between the time effects in certain ferro-magnetic and ferro-electric materials.

In the ferro-electric ceramics examined the time effects referred to were most pronounced in the barium-lead-titanate compounds, and increased with the lead content, while in the pure barium-titanate specimens the effect was practically non-existent. In this respect the phenomenon seems to be similar to the ferro-magnetic case, where it is also noted that the changes in the magnetization with time are also greater in the high-coercivity materials. Considerable time effects were also discovered in the single crystals used in the measurements.

The period of time over which significant changes in the polarization occurred also proved to be longer in the high coercivity specimens, although only in the 96 : 4 barium–lead–titanate ceramic did measurable increments continue for lengths of time comparable to those found in Dr. Street's experiments.

The increase of the polarization with time, subsequent to the application of a d.c. field to a ferro-electric, is most readily studied in two ways: (*a*) by recording the decay with time of the Barkhausen noise produced by the field, or (*b*) measuring the variation of the incremental permittivity. A limited number of results from these two methods were available and the examination of these is described below.

Ideally, to examine the increase of the polarization with time from a study of the Barkhausen effect, the noise voltage should be integrated continuously over the period of measurement. In the present work the r.m.s. noise level was recorded as a function of time. To assimilate the above as closely as possible the level was averaged over equal intervals of time, and these values, taken as being proportional to the increase in the polarization, were plotted against log t.

It was found that for values of t greater than about 1·5 sec the relationship was indeed linear for the time over which the record was taken—20–30 sec. By the more sensitive method of counting the number of Barkhausen pulses produced by the applied stress, the time effects could be studied over longer periods. The analysis of such a record for the 98 : 2 barium–lead–titanate showed an essentially linear relationship between pulse number (i.e. ΔP) and log t, over a period of 3 min, but then fell away significantly.

A similar examination of the incremental capacitance-time results revealed the same behaviour. With the single crystal and the 98 : 2 barium–lead–titanate specimen the time of measurement was some 60 sec, and in each case the decrease of this quantity (which is proportional to the increase in polarization), plotted against log t, showed a linear relationship. The same linearity was observed in the case of the 96 : 4 barium–lead specimen, and in this instance held over the 240 sec of measurement.

It should be pointed out that in all these cases the applied stress was increased from zero and was almost invariably greater than the coercive field of the specimen. No results are therefore available on the time effects at various points on the hysteresis cycle.

Likewise no results are available which might afford quantitative information on the effect of temperature on the phenomenon. In this connection, however, the reference made below to work by Merz seems pertinent.

The fact that the polarization established by an applied field depended on its time of application raised the question whether there would be any decrease in the maximum polarization produced by a.c. fields of increasing frequency. An investigation of this was carried out on the crystal used in the noise investigations. The strength of the applied field was kept constant and its frequency increased from 50 c/s to 20 kc/s. The resulting hysteresis loops were photographed and the values of the maximum polarization measured. When these were plotted against the logarithm of the frequency a linear decrease in polarization was revealed up to about 15 kc/s, when a sharper fall-off became apparent. Over a smaller frequency range a similar result was obtained from the measurement of the capacitance of the specimen on a high-voltage Schering bridge.

It is possible to criticize these results on the basis that the effect may have been due to a temperature rise brought about by the large internal heating associated with the hysteresis losses. However, measurements of the actual rise in temperature of the specimen indicated that this was not sufficient to produce the effects associated with the upper transition point of the material. Moreover, forced cooling did not influence the results significantly. It is felt that the effect was quite genuine.

It is not suggested, however, that this behaviour is common to all crystals of $BaTiO_3$, but will be determined to a large extent by the physical form of the crystal and any strains, etc., present in it. Similar effects unique to individual specimens have been reported in connection with relaxations in certain ferrites.[A,B]

This "lag" is also referred to by Merz,[C] who has shown that the strain components of the basic lattice structure of $BaTiO_3$ should be proportional to the square of the spontaneous polarization in the same way as is found in other ferro-electrics.

However, this relationship is confirmed only from the Curie point down to about 90° C, while at lower temperatures the measured values of the spontaneous polarization are smaller than expected. It is found that this discrepancy decreases with increasing field strength and temperature annealing. Merz concludes that the difference is, at least in part, due to the existence of anti-parallel domains which persist long enough at 50 c/s at low temperatures to prevent the crystal from polarizing completely. The influence of temperature suggests that the effects depend to a large extent on the physical state of the specimen.

It is tempting to conclude that this frequency dependence of the polarization and the long-time effects following the application of a direct voltage, are manifestations of the same phenomenon, but the evidence is slight.

Other time effects have been reported in the titanate group. In work on piezo-electric constants Bradfield[D] noted a creep under steady fields, which continued for several hours, and he suggested that some process with a relaxation time of several minutes was taking place. A comparison with the course of the incremental capacitance variation showed that this proceeded about six times faster.

The whole problem of the time effects in $BaTiO_3$ is obviously fairly complicated, but what results are available suggest that some of the effects at least may be explained in terms of Dr. Street's "ferro-electric viscosity."

Additional References

(A) Went, J. J., and Wijn, H. P.: *Physica*, 1951, **17**, p. 976.
(B) Galt, J. K.: *Physical Review*, 1952, **85**, p. 644.
(C) Merz, W. J.: M.I.T. (Laboratory for Insulation Research) Technical Report No. 19, June, 1949.
(D) Bradfield, G.: National Physical Laboratory Report: Vol. 7, Series 9, 1950.

621.317.784.029.6

Paper No. 1888 R
Sept. 1955

A MILLIWATTMETER FOR CENTIMETRE WAVELENGTHS

By A. C. GORDON-SMITH.

(*The paper was first received 2nd March, and in revised form 5th May, 1955.*)

SUMMARY

The instrument is in the form of a differential air thermometer, and consists of two similar glass cells connected by a glass capillary tube. A carbon-coated strip is placed in each of the two cells; one of these strips absorbs the radio-frequency power to be measured and the other is heated by direct current. A liquid pellet in the capillary tube is displaced as a result of the differential expansion of the air in the two cells arising from the heat dissipated in the carbon-coated strips. After balancing the system by adjustment of the d.c. cell the radio-frequency power is finally determined by the measurement of the d.c. power which must be substituted for the radio-frequency power in the same cell in order to maintain the state of balance. The instrument has been used to measure powers of 10 to 100 mW at a wavelength of 3 cm with a discrimination of about 0·2 mW.

(1) INTRODUCTION

Until recently, centimetre-wave powers of the order of milliwatts have been measured by some form of bolometer or thermistor in a bridge circuit. It is known that these instruments do not respond in the same manner to both direct-current and radio-frequency heating, particularly as the wavelength decreases through the centimetre band; and they are normally calibrated by reference to a water-calorimeter,[1,2] though Cullen[3] has now demonstrated the possibility of another standard which depends upon the torque exerted on a vane suspended in a waveguide through which the radio-frequency energy to be measured is passing. For accurate measurements, both the water-calorimeter and Cullen's wattmeter require a power of 10 watts or more. This means that a known attenuator must be used to calibrate a thermistor at the milliwatt level in terms of such standards.

A considerable number of precautions have to be taken in the calorimetric method of power measurement if the error is not to exceed 2%, and some of these have been discussed in detail by Bailey, French and Lane.[2] It would clearly be desirable to have an instrument capable of direct measurement of powers in the range 10–100 mW, thus avoiding calibration by comparison with a power a hundred times or more greater and the use of an attenuator, which may add a further element of uncertainty. It is believed that the instrument described in the paper will meet this need.

(2) DESCRIPTION

The milliwattmeter, illustrated diagramatically in Fig. 1, consists essentially of a differential air thermometer operating in a manner first described by Fleming.[4] Two similar glass cells are connected together by a piece of capillary tubing containing a liquid pellet which serves to indicate the differential expansion of the heated air in the two cells. Each cell contains a carbon-coated strip, one to absorb the radio-frequency power, the other being heated by direct current. The two cells and carbon strips are made as nearly identical as possible in order to compensate for ambient temperature fluctuations. The following procedure is adopted in making a measurement. The expansion of the air in the one cell, caused by the heat communicated from the carbon strip absorbing the radio-frequency power, is counterbalanced by the expansion of the air in the other cell produced by a corresponding d.c. power applied to its carbon strip. In this balancing process the liquid pellet is kept stationary, and there is thus no actual expansion in either cell but only an increase in pressure. During tests on the instrument the equality of the two cells was checked by applying d.c. heating, and under balance conditions the d.c. powers differed by only about 2 or 3%. In order to allow for the slight differences in the characteristics of the two cells the magnitude of the radio-frequency power is finally obtained by determining what d.c. power must be substituted for the radio-frequency power in the same cell to maintain the state of balance.

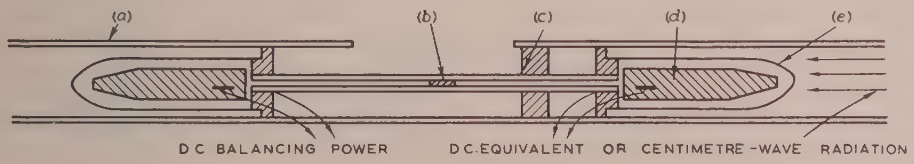


Fig. 1.—Milliwattmeter.

(*a*) Waveguide.
(*b*) Liquid pellet in capillary tube.
(*c*) Metal plate.
(*d*) Carbon strip.
(*e*) Glass cell.

The air cells are made of glass tubing with an outside diameter of 1 cm and a wall thickness of 0·05 cm, and they are tapered at one end as shown in Fig. 2; the overall length of each cell is 5 cm. The internal diameter of the capillary tube connecting the cells is 0·5 mm and its length is about 30 cm. The absorbing elements are made of cardboard strip coated with carbon, and they are bent into the form of a V with the carbon on the inside. They are tapered as shown in Fig. 2 in such a manner that the heating in each case leads to similar temperature gradients along them. In addition, it is found that this taper, in conjunction

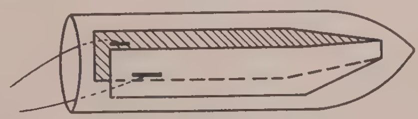


Fig. 2.—Glass cell and carbon strip in perspective.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

The paper is an official communication from the Radio Research Station, Department of Scientific and Industrial Research.

with an adjustable metal back-plate, gives a very good matched termination to the waveguide along which the radio-frequency power is flowing. The metal plate reflects the small fraction of power, which would otherwise be lost, back into the absorbing strip. Various liquids have been tried for the indicating pellet and a silicone oil of low viscosity has finally been selected as the most suitable.

So far, the milliwattmeter has been used only at a wavelength of 3 cm. It is mounted in a section of rectangular waveguide of the same dimensions as that along which the power to be measured is transmitted, and an aperture is cut in the upper face of the waveguide so that the movement of the indicating pellet may be observed. When the highest attainable sensitivity is required a low-power microscope is used.

(3) PERFORMANCE

Measurements have shown that the voltage standing-wave ratio achieved when the milliwattmeter is included in the waveguide system as described above is not worse than 0·9, so that no additional impedance transformation is required.

If d.c. or radio-frequency power is applied to one cell, the other remaining unheated, it is found that the deflection of the pellet varies linearly with input power over the range 10 to 100 mW and is at the rate of 1·8 mm/mW. This indicates that the rise in the mean temperature of the air in the cell varies linearly with input power and implies that all the heat lost from the cell to the surroundings is also a linear function of the input power. It would thus appear that the rise in the mean temperature of air inside the cell would not be influenced by the nature of the temperature gradient along the absorbing strip.

The following is a detailed description of the manner in which a measurement is carried out. Between the radio-frequency source and the milliwattmeter there is inserted a flap-attenuator by means of which the attenuation can be varied rapidly between zero and 30 dB. The attenuator is operated in conjunction with a switch in the d.c. supply to the radio-frequency cell so that it is possible to effect a quick substitution of d.c. for radio-frequency power. The indicating pellet is brought to a balanced position with radio-frequency power in one cell and d.c. power in the other. Owing to the thermal inertia of the system the establishment of a balance under steady-state conditions takes several minutes, but the final adjustments, involving only small increments of power, may be made much more rapidly. The balance condition is repeated with d.c. power substituted for the radio-frequency power, and it is ascertained that the balance is not disturbed by switching between the two forms of heating. This procedure retains the advantage of steady-state conditions whilst minimizing the effect of fluctuation in the ambient temperature. The d.c. power equivalent to the radio-frequency power is then determined with the aid of a voltmeter and milliammeter, both of which must be accurately calibrated.

Further evidence that the heat is dissipated in the absorbing element in a similar manner for both kinds of heating is found from the fact that when d.c. is substituted for radio-frequency power under balance conditions at various levels the initial balance in each case remains unaltered apart from small fluctuations that may arise from ambient temperature changes.

From subsidiary measurements on the absorbing properties of the glass cell, it is estimated that about 1% of the radio-frequency power to be measured may be absorbed by the cell, and although the heating so caused would contribute to the final increase in the temperature of the air inside the cell, there would be a small residual error in the determination of the power, since a similar effect is not present when the cell is heated by direct current. This error could be virtually eliminated if the cells were made of a suitable low-loss material.

(4) COMPARISON WITH OTHER METHODS OF POWER MEASUREMENT

With the aid of a directional coupler having a coupling factor of $15\cdot50 \pm 0\cdot05$ dB a power of about 40 mW was measured by means of the milliwattmeter described, and the power at a level 15 dB lower was determined simultaneously by a thermistor bridge, which had previously been calibrated by reference to a water-calorimeter in the manner described by Bailey, French and Lane.[2] The agreement between the two measurements, allowing for the coupling factor, was well within the limits of experimental error, estimated to be about 2%.

(5) CONCLUSIONS

It is considered that the milliwattmeter described in the paper is a convenient and reliable instrument suitable for the direct measurement of radio-frequency power in the range 10 to 100 mW at wavelengths of about 3 cm. There seems to be no reason why the instrument should not be designed for both longer and shorter wavelengths in the centimetre and millimetre region.

Further work at present in progress, in which metallic films on mica strips are used as the absorbing elements, indicates that it should be possible to achieve greater sensitivity and so to measure powers much less than 10 mW.

The work described has been concerned mainly with the production of an accurate instrument suitable for use in the laboratory. It is possible, however, that an application might be found for a simpler construction in which no d.c. heating was necessary and in which the displacement of the pellet was calibrated directly in terms of radio-frequency power.

(6) ACKNOWLEDGMENTS

The author wishes to acknowledge the advice given by Dr. J. A. Saxton in the course of the work described, which was carried out as part of the programme of the Radio Research Board. The paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

(7) REFERENCES

(1) STREET, R., and WHITAKER, P. D.: "The Measurement of Microwave Power at Wavelengths of 3 cm and 10 cm," *Proceedings of the Physical Society*, 1950, **63B**, p. 623.

(2) BAILEY, R. A., FRENCH, H. A., and LANE, J. A.: "The Comparison and Calibration of Power-Measuring Equipment at Wavelengths of 3 cm and 10 cm." *Proceedings I.E.E.*, Paper No. 1689 R, September, 1954 (**101**, Part III, p. 325).

(3) CULLEN, A. L., and STEPHENSON, I. M.: "A Torque-Operated Wattmeter for 3-cm Microwaves," *ibid.*, 1952, **99**, Part IV, p. 294.

(4) FLEMING, J. A.: "Some Quantitative Measurements in Connection with Radiotelegraphy," *Journal I.E.E.*, 1910, **44**, p. 344.

621.396.677.3 : 621.397.6

Paper No. 1890 R
Sept. 1955

V.H.F. AERIALS FOR TELEVISION BROADCASTING

By G. J. PHILLIPS, M.A., Ph.D., B.Sc., Associate Member.

(*The paper was first received* 2nd *April, and in revised form* 10*th May*, 1955.)

SUMMARY

The "batwing," which forms the radiating element of the R.C.A. Superturnstile aerial, can be used to take the place of wide-band dipoles in certain other aerial designs. The paper describes briefly some examples for both horizontally and vertically polarized transmissions; unlike the Superturnstile they are suitable for masts of thickness comparable with a wavelength. The immediate application is to television broadcasting in the United Kingdom in both Band I (41–68 Mc/s) and Band III (174–216 Mc/s). Another feature of new designs of transmitting aerial is the use of a split system having separately fed halves to avoid the need for a reserve aerial.

(1) GENERAL CONSIDERATIONS

It is becoming general practice for a number of broadcasting services to be radiated from the same site, by the use of a single high mast. For example, some masts at British Broadcasting Corporation stations may, in the near future, be carrying aerials for transmitting programmes in Bands I (41–68 Mc/s), II (88–95 Mc/s) and III (174–216 Mc/s). The aerial systems for frequencies in the different bands are preferably arranged one above the other rather than interlaced, in order to simplify both design and maintenance. Since each system will have more than one tier of elements to provide sufficient aerial gain, the mast must be of large section—except perhaps at the top—to give the required mechanical strength. It is then inevitable that some of the aerial systems will be mounted on a mast whose thickness is comparable with a wavelength. Thus, at many future B.B.C. stations a square-section mast of about 4 ft side will be used.

For thin masts, satisfactory aerial designs are well known; these include, for horizontally polarized transmission, the R.C.A. Superturnstile aerial[1] which uses radial elements, and for vertically polarized transmission the four-dipole system used at Sutton Coldfield[2] and other high-power British stations. In each case the four elements of one tier are fed in phase rotation. On square-section masts of side greater than about $0 \cdot 15\,\lambda$ these systems no longer give a horizontal radiation pattern that is sufficiently uniform for general broadcasting purposes.

For thicker masts an arrangement consisting of a ring of in-phase dipoles (horizontal or vertical according to the polarization required) can usually be found for which the horizontal radiation pattern is satisfactory, though the number of dipoles required may sometimes be inconveniently large. In many systems of this kind, pairs of dipoles may be replaced by a wide-band batwing element as illustrated in Fig. 1, without appreciably affecting the radiation pattern. This follows the line of development of the Superturnstile from the simple turnstile, whereby two tiers of the simple turnstile are replaced by a single tier of batwing elements; thus in the Superturnstile the batwings occupy two planes at right-angles and have a central support pole as seen in the plan view, Fig. 2(*a*). But for general application the batwing unit may also be used without a central pole, as illustrated in Fig. 1(*b*).

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Dr. Phillips is with the British Broadcasting Corporation.

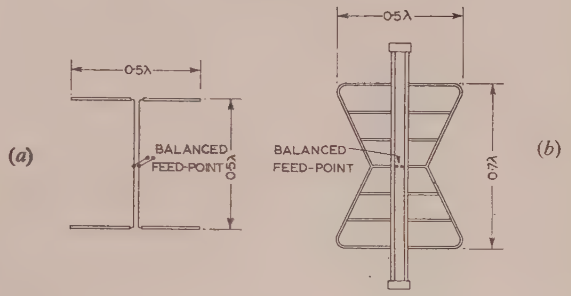


Fig. 1.—Two in-phase dipoles, (*a*), may be replaced by a batwing element, (*b*), having a closely similar radiation pattern.
Approximate dimensions are in terms of mid-band wavelength, λ.

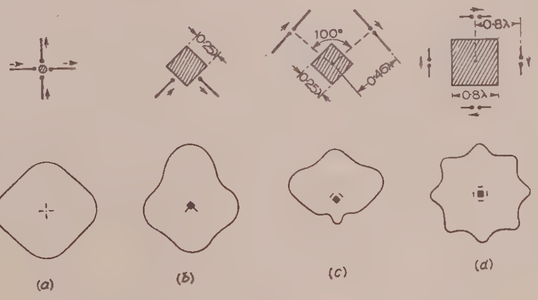


Fig. 2.—Systems for horizontal polarization.
Upper Figures are plan views. → In-phase radiating currents. → Radiating currents in phase quadrature.
Lower Figures are horizontal-radiation patterns from relative-field-strength measurements with models at 450 Mc/s.

The advantages of using batwings instead of dipoles are as follows:

(*a*) The bandwidth of the batwing unit is at least as good as that of a wide-band dipole which has a parallel stub to give optimum susceptance compensation.

(*b*) The number of feed points of the complete aerial system is reduced.

(*c*) Mechanical support is generally more convenient.

(*d*) A wide choice of impedance (60 to 150 ohms) is available by suitable adjustment of dimensions.

The descriptions given in the paper are of systems in which batwings may be used, and which may be mounted on thick masts. Besides systems for horizontal polarization in which batwings reduce the number of tiers required—as in the Superturnstile—there are systems for vertical polarization in which the number of elements per tier is halved by replacing dipoles by batwings. Although in the application to television broadcasting the advantages of batwings are sufficiently attractive to warrant incorporating them in the design, there may be other applications, perhaps not requiring a wide band, in which the dipole system from which they were derived might be preferred.

(2) HORIZONTALLY POLARIZED TRANSMISSION

Fig. 2 shows a plan view of aerial systems for radiating horizontally polarized waves, together with their horizontal radiation patterns. The Superturnstile is shown in Fig. 2(*a*). The radiation pattern illustrated assumes a relatively thin supporting pole, and represents a $\pm 0 \cdot 5$ dB variation, but it must be remembered that the variation increases rapidly with size of pole, reaching about $\pm 1 \cdot 5$ dB for a pole of $0 \cdot 2\lambda$ diameter.

Fig. 2(*b*) shows a variant of the radial-element system. Two batwings are mounted perpendicular to the faces of a square-section mast and fed in antiphase. The batwing on each face is one-half of that illustrated in Fig. 1(*b*), since the parent system consists of quarter-wave rods projecting from the mast—an adaptation of the V- or quadrant-aerial.[4] If the mast-side dimension is about a quarter-wavelength the horizontal radiation pattern is nearly uniform, as illustrated. Some of the proposed low-power B.B.C. stations will radiate horizontally polarized waves in Band I, and a system based on the idealized arrangement of Fig. 2(*b*) will be used on a 4 ft-square mast at most of these.

The systems shown in Figs. 2(*c*) and 2(*d*) are tangential arrangements of batwings, i.e. the units [Fig. 1(*b*)] are mounted away from the mast and tangential to a cylinder centred on the mast axis. The system shown in Fig. 2(*c*) is also for a mast of about a quarter-wavelength side but is an arrangement deliberately chosen to give a directional pattern. This form of aerial was developed for the Band I transmission at the Norwich station, because the power that could be radiated over a wide sector to the south had to be limited to avoid interference in parts of southern England and on the Continent, where the same channel is being used.

A system with the dimensions shown in Fig. 2(*d*) gives a fairly uniform horizontal radiation pattern and is suitable for Band III transmissions with a square-section mast of 4 ft side. Such an arrangement of four tangential batwings, with suitable adjustment of the spacing from the mast, can be made to give a fairly uniform pattern for a square-section mast whose side is in the range $0 \cdot 5$ to $0 \cdot 9\lambda$, or less than about $0 \cdot 24\lambda$.

(3) VERTICALLY POLARIZED TRANSMISSION

The batwing unit of Fig. 1(*b*) can be used with the slot aperture horizontal in arrays for radiating vertically polarized waves. Fig. 3 shows a provisional design for one tier of a Band III aerial suitable for the cylindrical portion of a mast such as is used at most high-power B.B.C. stations. Four batwings, mounted tangentially round the 6 ft 6 in-diameter mast as shown, and fed in phase, can give a reasonably uniform horizontal radiation pattern even though the mast diameter exceeds a wavelength in Band III; the variation with bearing is between $+2 \cdot 5$ dB and $-2 \cdot 2$ dB relative to the mean.

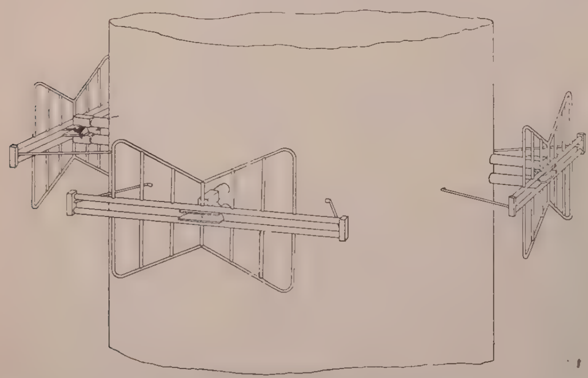


Fig. 3.—Provisional arrangement for vertically polarized Band III aerial on a 6 ft 6 in-diameter cylinder.

Four batwings mounted in the same way on a square-section mast of 4 ft side would give a considerably more uniform horizontal radiation pattern. Comparison with the system of Fig. 2(*d*) for horizontal polarization is interesting, since, in the event of Band III transmissions being required at a later date from the newer sites of the B.B.C., satisfactory designs for either kind of polarization—similar in basic requirements for elements and feeder system—would be available for mounting on the 4 ft-side masts.

It should be emphasized here that the reduction in the number of feed points by the use of batwing elements in place of dipoles, in systems for horizontal polarization, is achieved because the number of tiers required to fill a given vertical aperture is reduced. But the factor of reduction is by no means always as low as one-half in systems of this type, though there is usually a worth-while improvement. Systems for vertical polarization can achieve the full reduction of one-half in the number of feed points within each tier, the optimum tier spacing being unaffected; but, unless there is a satisfactory parent system with dipoles $\frac{1}{2}\lambda$ apart, the use of one batwing for every two dipoles may seriously degrade the horizontal radiation pattern.

(4) OTHER DEVELOPMENTS

A new feature of most television aerial systems to be installed at future B.B.C. stations is that they will be split into upper and lower halves. Each half will be fed through a separate transmission line running up the mast, and the halves will normally radiate in phase. In the event of a fault in one half the other half may continue to be used. With full power in one half of the aerial the effective radiated power would be reduced below normal by only about 3 dB. This arrangement allows a separate reserve aerial to be dispensed with.

Future developments in television may require the transmission of two Band III programmes simultaneously from the same site, in which case there are obvious advantages in using a single Band III aerial array for the purpose. Batwing arrays of the type described can be matched very well to the feeder system (e.g. to give a standing-wave ratio better than $0 \cdot 9$) over the whole of Band III. With two channels spaced 15 or 20 Mc/s apart there should be no great difficulty either in combining both pairs of sound and picture signals into one feeder system or in trimming the aerial end of the feeder system to give the high degree of matching required at the two vision carrier frequencies.

(5) ACKNOWLEDGMENTS

The author is indebted to the Chief Engineer of the British Broadcasting Corporation for permission to publish the paper. The designs described are some of those being developed and engineered in the Research Department and the Planning and Installation Department of the Corporation.

(6) REFERENCES

(1) WOLF, F. J.: "Triplex Antenna for Television and F.M.," *Electronics*, July, 1947, p. 88.
(2) BEVAN, P. A. T., and PAGE, H.: "The Sutton Coldfield Television Broadcasting Station," *Proceedings I.E.E.*, Paper No. 1116 R, November, 1951 (**98**, Part III, p. 416).
(3) CARTER, P. S.: "Antenna Arrays Around Cylinders," *Proceedings of the Institute of Radio Engineers*, 1943, **31**, p. 671.
(4) WELLS, N.: "The Quadrant Aerial: an Omni-Directional Wide-Band Horizontal Aerial for Short Waves," *Journal I.E.E.*, 1944, **91**, Part III, p. 182.

621.375.221.2

Paper No. 1889 R
Sept. 1955

ON DISTRIBUTED AMPLIFICATION

By D. G. SARMA, M.Sc.

(*The paper was first received* 27*th September*, 1954, *and in revised form* 19*th April*, 1955.)

SUMMARY

The response characteristics of a distributed amplifier are discussed. It is shown that they may be improved by making the delay characteristics of the anode line and the grid line different. This method, which may be called "staggering," may be applied with advantage to almost all the conventional circuits employed in constructing distributed amplifiers. It enables one to obtain a flatter gain/frequency characteristic and a more linear phase-shift/frequency characteristic than could be obtained with the corresponding unstaggered amplifier. Photographic records of the response of experimental amplifiers to a step-voltage input show this improvement very clearly.

(1) INTRODUCTION

The principle of distributed amplification was originally suggested by Percival and has since been developed by many other workers. A distributed amplifier consists mainly of two transmission lines of identical characteristics, called the grid line and the anode line, with a number of amplifying valves arranged between the two lines in the manner shown in Fig. 1. The lumped

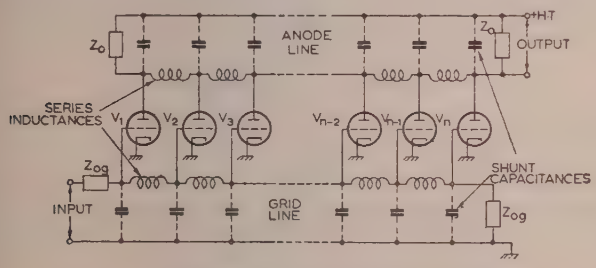


Fig. 1.—Circuit diagram of a distributed amplifier.

transmission lines are terminated properly at each end. A signal voltage applied at the input end of the grid line travels down the line to the other end where it is absorbed by the terminating impedance. This travelling signal is picked up by the several valves, the grids of which are connected to the grid line as shown. From the anode of each valve the amplified signals travel in both directions. Since the grid and anode lines have identical delay characteristics, the forward-travelling signals from each of the anodes appear at the output terminating impedance at the same time and add together to give an amplified output. The backward travelling signals, on the other hand, arrive at the corresponding termination at different instants, and the voltage at this termination of the anode line is a succession of input signals, amplified and spaced equally apart. This voltage is of no consequence for amplification, and is lost in the non-reflecting terminating impedance.

For faithful reproduction of the input signal an amplifier is required to have a large bandwidth with flat amplitude and delay characteristics. The lumped transmission lines used in a distributed amplifier consist of low-pass filter sections arranged in cascade and terminated at both ends by the proper termination impedances. The bandwidth of the amplifier as a whole depends on the cut-off frequency of the low-pass filter sections. The overall gain characteristic is governed by the variation of the characteristic impedance of the anode line within the pass band, and the overall phase-shift characteristic is identical with that of the lumped lines. Since the characteristic impedance of such lumped lines consisting of series-inductance and shunt-capacitance elements increases as the cut-off frequency is approached, the amplifiers show a peak in the gain characteristic below the cut-off frequency. Also, in general, the phase-shift characteristics of low-pass filter sections deviate from linearity as the cut-off frequency is approached. Attempts to flatten out the peak in the gain characteristic have given rise to a number of circuit arrangements. Similarly, attempts have been made to linearize the phase-shift characteristic by using special networks. Most of these methods replace the simple constant-*k* *LC* filter network by more complicated networks which exhibit flatter gain and more linear phase-shift characteristics. Thus mutual coupling between the coils forming the successive series elements of the line has proved to be useful in improving the phase-shift characteristic, and pairing of the anodes and introducing lossy elements in the filter networks have been used to improve the gain characteristic.[1,2]

A new method is described in the paper which improves both the gain and the phase-shift characteristics at the same time. This method may be termed "staggering." The lumped lines in the distributed amplifier are so arranged that the anode-line travelling wave and the grid-line travelling wave are not in phase at corresponding points along the lines. It may appear at first that this will produce distortion by spacing the component signals (i.e. the signals which appear at the output through the different valves) in time. But it is found that the steady-state gain and phase-shift characteristics may be improved considerably by properly adjusting the stagger. The principle underlying this improvement is quite simple and is explained in Section 2 with reference to the constant-*k* *LC* filter network.

(2) PRINCIPLE OF STAGGERING

In an amplifier embodying the constant-*k* *LC* filter network as the elements of the lumped lines, the stagger is introduced by making the cut-off frequency of the grid line a little higher than that of the anode line.

At a given frequency, a line with a higher cut-off frequency produces a smaller phase shift than one with a lower cut-off frequency. Also, the difference between the phase shifts produced by the two lines increases continuously as the frequency is increased.

With the above arrangement, therefore, the anode line will introduce a phase shift larger than that introduced by the grid line. In Fig. 1, let the phase shifts introduced by the individual sections of the anode line and the grid line be θ_p and θ_g respectively.

The improvement in the steady-state characteristic of the amplifier can be understood in the following way: Let OA in Fig. 2 represent the voltage vector at a particular frequency at

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. Sarma is at the Institute of Radio Physics and Electronics, University of Calcutta.

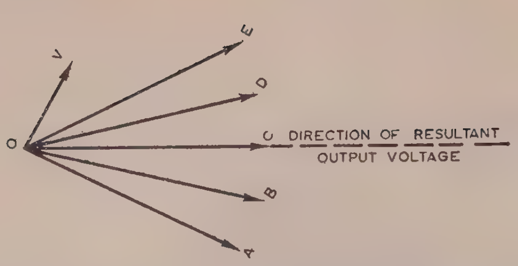


Fig. 2.—Vector diagram to illustrate improvement in phase-shift characteristic due to "staggering" in 5-valve amplifier.

OV = Input voltage.
OA, OB, OC, OD, OE = Output voltages due to valves V_1, V_2, V_3, V_4, V_5 respectively.

the output terminals of the amplifier due to the component signal which has travelled through the first valve and then through the anode line. Its phase angle with respect to the input voltage is $\pi + (n - 1)\theta_p$, n being the number of valves in the amplifier. The component signal, which travels along the first section of the grid line and then arrives at the output terminals through the second valve and the remaining portion of the anode line, suffers a phase shift of $\theta_g + (n - 2)\theta_p + \pi$, and consequently lags behind the voltage OA by an angle $(\theta_p - \theta_g)$. Let this voltage be represented by the vector OB. In a similar manner, the component signal which arrives at the output through the third valve travels through one section more of the grid line and one section less of the anode line, and consequently falls behind OB by the same phase angle $(\theta_p - \theta_g)$. Thus, the overall output, which is the vector sum of these and all other such component signals travelling through different portions of the anode and grid lines, suffers a net phase shift which is less than that introduced by the anode line from the anode of the first valve to that of the last valve, and greater than the phase shift introduced by the grid line between the grids of the first and last valves. As is shown later, the phase shift of the final output is, in fact, the average of these two phase shifts. Since the grid-line cut-off frequency is made higher, the phase-shift characteristic of the grid line is such that its deviation from linearity is less than that of the phase-shift characteristic of the anode line at all frequencies below the anode-line cut-off frequency. Consequently, the average of the two characteristics is more linear than the phase-shift characteristic of the anode line (see Section 10.1). That is, the linearity of the overall phase-shift characteristic of the amplifier is better than would have been obtained by making the cut-off frequencies of the two lines identical and equal to that of the anode line (see Fig. 3).

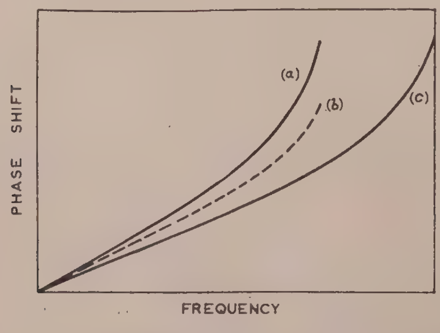


Fig. 3.—Curves illustrating the principle of phase-shift improvement by staggering.

(a) Phase-shift characteristic of anode line.
(b) Overall phase-shift characteristic of amplifier.
(c) Phase-shift characteristic of grid line.

The improvement in the gain characteristics is easily understood. If the stagger is adjusted in such a manner that the vectors OA, OB, etc., arrange themselves so that the resultant is zero at some frequency just below the cut-off, the peak in the gain characteristic can be avoided.

Tischer[3] has shown that the gain-bandwidth product of an amplifier is approximately proportional to

$$\int_0^\infty A(\omega) \cos [\Delta\phi(\omega)]d\omega$$

where $A(\omega)$ is the amplitude characteristic and $\Delta\phi(\omega)$ is the departure of the phase characteristic from linearity. When a distributed amplifier is staggered, both $A(\omega)$ and $\Delta\phi(\omega)$ are modified. Therefore, the gain-bandwidth product will improve if the reduction in $\int_0^\infty A(\omega)d\omega$ is more than compensated by the reduction in $\Delta\phi(\omega)$. Photographic records of the transient response of experimental amplifiers (Figs. 14 and 15) show that there is a slight overall improvement in the rise time of the staggered amplifiers. Since the gain is the same in both amplifiers, this means a corresponding improvement in the gain-bandwidth product.

We shall now proceed to study the effect of staggering on the gain and phase-shift characteristics of distributed amplifiers with three different types of network.

(3) AMPLIFIER WITH CONSTANT-*k* *LC* FILTER NETWORK

The overall gain characteristic of a staggered distributed amplifier with constant-*k* *LC* filter network is given (see Section 10.2) by

$$A(x) = g_m \frac{Z_{0\pi}}{2} \frac{\sin n\psi/2}{\sin \psi/2} \quad . \quad . \quad . \quad . \quad (1)$$

and the phase shift by

$$\phi(x) = (n - 1) (\text{arc} \sin x + \text{arc} \sin qx) \quad . \quad . \quad . \quad (2)$$

where g_m = mutual conductance of the valves employed.
$Z_{0\pi}$ = characteristic impedance of the anode line $= \sqrt{\left(\frac{L}{C}\right)} \frac{1}{\sqrt{(1 - x^2)}}$
n = number of valves in the amplifier.
$\psi = (\theta_p - \theta_g) = 2 (\text{arc} \sin x - \text{arc} \sin qx)$.
ω_{0p} and ω_{0g} = anode- and grid-line cut-off frequencies respectively.
$q = \omega_{0p}/\omega_{0g}$.
x = normalized frequency $= \omega/\omega_{0p}$.

This gain is measured from the grid of the first valve to the anode of the final valve. The gain and phase-shift characteristics of the corresponding unstaggered amplifier are given by

$$A(x) = ng_m \frac{Z_{0\pi}}{2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

and

$$\phi(x) = 2(n - 1) \text{arc} \sin x \quad . \quad . \quad . \quad . \quad (4)$$

Curves of the gain and phase-shift characteristics for both the staggered and unstaggered cases are shown in Figs. 4 and 5 for $n = 5$ and for different values of q, the staggering factor.

In order that the gain may fall to zero just below the cut-off frequency of the anode line, the relation

$$q = \sin (\pi/2 - \pi/n) \quad . \quad . \quad . \quad . \quad . \quad (5)$$

has to be satisfied (see Section 10.2). This, however, is not the optimum value of q. From the curves in Figs. 4 and 5 it appears

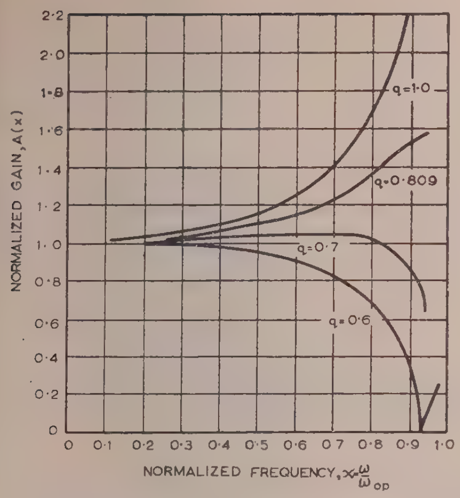


Fig. 4.—Normalized gain/frequency characteristic for different values of staggering factor q.

$q = 1$ corresponds to the unstaggered case.

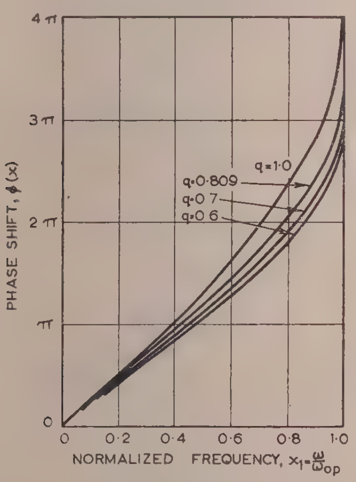


Fig. 5.—Normalized phase-shift/frequency characteristic for different values of staggering factor q.

$q = 1$ corresponds to the unstaggered case.

that the optimum value of q lies near $0 \cdot 7$ for $n = 5$, although eqn. (5) yields $q = \sin 54° = 0 \cdot 809$.

(4) AMPLIFIER WITH LINES HAVING MUTUAL COUPLING BETWEEN ADJACENT COILS

It is known that the phase-shift characteristic of an m-derived low-pass filter can be made linear up to about 60% of the pass band by making the value of m greater than unity. Such values of m require a negative inductance element in the shunt branch of the filter. This may be realized in effect by introducing a mutual coupling between the adjacent coils of the series elements of the line.

The delay characteristics of m-derived filter sections with m greater than unity are shown in Fig. 6. From this set of curves it is seen that the value of m for the best linearity of the phase-shift characteristic (i.e. for the flattest delay characteristic) is about $1 \cdot 27$. Thus, if a distributed amplifier is constructed with transmission lines consisting of m-derived filter sections with

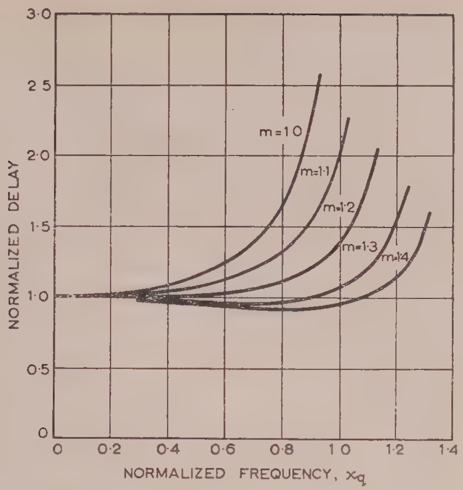


Fig. 6.—Normalized delay/frequency characteristic for different values of m.

$m = 1 \cdot 27$, its phase-shift characteristic will be linear up to about 60% of the pass band. Also, the cut-off frequency of such an m-derived filter section is higher than that of the corresponding constant-k filter ($m = 1$) by the factor m. This increase in the bandwidth, combined with the linearization of the phase-shift characteristic, results in an overall improvement of the amplifier response.

The phase-shift characteristic, as well as the gain characteristic, of a distributed amplifier with m-derived filter sections as mentioned above may be improved further by staggering. The principle underlying the improvement in this case is the same as that in the previous case, described in Section 3.

The stagger in this case is introduced by making the m-values of the anode line and the grid line different, the grid-line having the larger m.

Since the overall delay characteristic of a staggered distributed amplifier is the average of the delay characteristics of the anode and the grid line (see Section 10.1), the m-values of the lines are so chosen that the average of the two delay characteristics may be as flat as possible. A study of Fig. 6 reveals that there are a number of possible combinations which will satisfy this condition.

The gain characteristic of an m-derived staggered distributed amplifier is given (see Section 10.3) by

$$A(x_q) = \frac{g_m Z'_0}{2} \frac{\sin n\psi/2}{\sin \psi/2} \quad . \quad . \quad . \quad . \quad . \quad (6)$$

and the phase-shift characteristic by

$$\phi(x_q) = \frac{n}{2}(\theta_p + \theta_g) \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where x_q = normalized frequency $= \frac{\omega}{2} R_{0p} C_{pk} = \frac{\omega}{2} R_{0g} C_{gk} = m_p x_p = m_g x_g$.

m_p and m_g = m-values of the anode and grid lines respectively.

C_{pk} = anode–cathode (anode–earth) capacitance.

C_{gk} = grid–cathode capacitance.

R_{0p}, R_{0g} = anode- and grid-line characteristic impedances, respectively, at zero frequency.

θ_p = phase shift of the anode line per section $= 2 \arc\tan \frac{m_p x_p}{\sqrt{(1 - x_p^2)}}$.

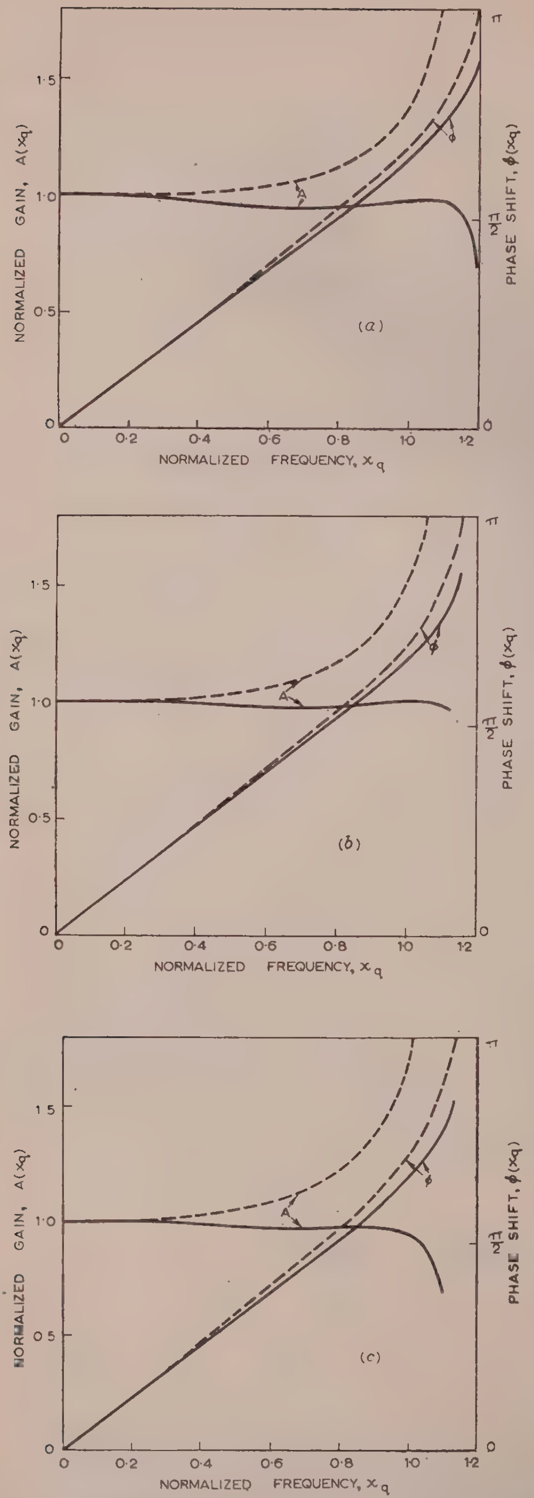


Fig. 7.—Normalized gain and phase-shift curves for 5-valve amplifier.

(a) —— Staggered: $m_g = 1{\cdot}4$; $m_p = 1{\cdot}2$.
---- Unstaggered: $m_g = m_p = 1{\cdot}2$.
(b) —— Staggered: $m_g = 1{\cdot}375$; $m_p = 1{\cdot}175$.
---- Unstaggered: $m_g = m_p = 1{\cdot}2$.
(c) —— Staggered: $m_g = 1{\cdot}4$; $m_p = 1{\cdot}14$.
---- Unstaggered: $m_g = m_p = 1{\cdot}14$.

$$\theta_g = \text{phase shift of the grid line per section} = 2 \arctan \frac{m_g x_g}{\sqrt{(1 - x_g^2)}}.$$

$$Z_0' = \frac{R_{0p} m_g}{\sqrt{(m_g^2 - x_q^2)}\sqrt{[1 + (1 - 1/m_g^2)x_q^2]}\sqrt{[1 + (1 - 1/m_p^2)x_q^2]}}.$$

$$\psi = \theta_p - \theta_g.$$

Curves of the gain and phase-shift characteristics for three combinations of m-values are shown in Figs. 7(*a*), 7(*b*), 7(*c*). In Fig. 7(*a*), $m_g = 1{\cdot}4$ and $m_p = 1{\cdot}2$; in Fig. 7(*b*), $m_g = 1{\cdot}375$ and $m_p = 1{\cdot}175$; and in Fig. 7(*c*), $m_g = 1{\cdot}4$ and $m_p = 1{\cdot}14$. The number of valves taken is 5. The value of $R_{0g}C_{gk}$ is kept equal to $R_{0p}C_{pk}$. This means that the ratio of the cut-off frequencies of the anode and grid lines is made equal to the ratio of their m-values.

The gain and phase-shift characteristics of the corresponding unstaggered distributed amplifier (obtained by putting $m_g = m_p = m$) are

$$A_0(x_q) = n\frac{g_m R_{0p}}{2} \frac{m}{\sqrt{(m^2 - x_q^2)}\left[1 + \left(1 - \frac{1}{m^2}\right)x_q^2\right]} \quad . \quad (8)$$

$$\phi_0(x_q) = 2n \arctan \frac{m x_q}{\sqrt{(m^2 - x_q^2)}} \quad . \quad . \quad . \quad (9)$$

These are also plotted in Figs. 7(*a*), 7(*b*) and 7(*c*) in dotted lines. The improvement in each case is noteworthy.

(5) AMPLIFIER WITH BRIDGED T-NETWORK

The bridged T-network shown in Fig. 8 may also be used in building distributed amplifiers.[1] If this network is adjusted so that

$$C_2(1 + \alpha) = 4C_1 \quad . \quad . \quad . \quad . \quad . \quad (10)$$

it behaves as an all-pass line with constant characteristic impedance independent of frequency.

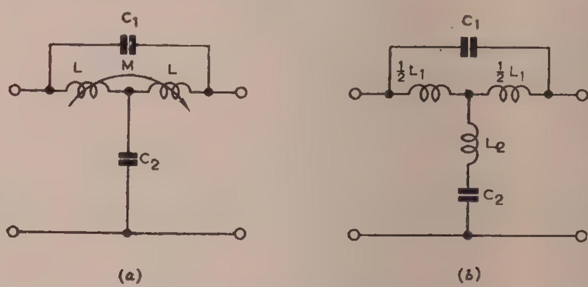


Fig. 8.—Bridged T-network: equivalent circuits.

(*a*) Actual.
(*b*) Theoretical, in which $L_2 = -M$, and $\alpha = 4L_2/L_1$.

The gain and phase-shift characteristics of an unstaggered amplifier under these conditions are given[1] by

$$A(x_q) = n\frac{g_m R_l}{2} \frac{1}{[1 - x_q^2(1 + \alpha)]^2 + x_q^2} \quad . \quad . \quad (11)$$

and

$$\phi(x_q) = 2n \arctan \frac{x_q}{1 - x_q^2(1 + \alpha)} \quad . \quad . \quad . \quad (12)$$

where $x_q = \frac{\omega}{2}R_l C_{pk} = \frac{\omega}{2}R_l C_{gk}$.

R_l = characteristic impedance of the lines.

$$1 + \alpha = 1 + 4\frac{L_2}{L_1} = \frac{1 - k}{1 + k}.$$

k = coefficient of coupling between adjacent coils.

The other symbols have the same meaning as before. The delay is given by

$$\tau = \frac{2}{\omega_c} \frac{1 + x_q^2(1+\alpha)}{[1 - x_q^2(1+\alpha)]^2 + x_q^2} \text{ seconds} \quad . \quad . \quad (13)$$

where $\omega_c = \dfrac{2}{R_l C_{pk}}$

The gain and delay characteristics are plotted in Figs. 9 and 10.

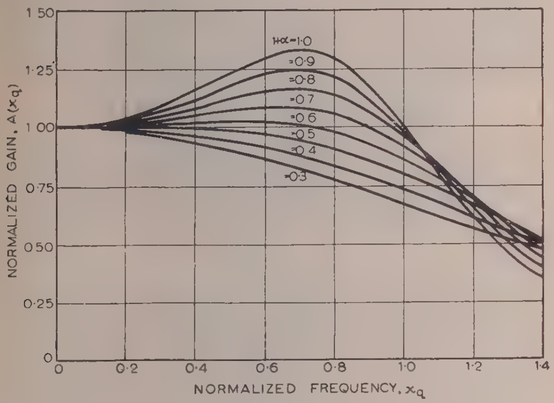


Fig. 9.—Normalized gain/frequency curve for bridged T-network for different values of $(1 + \alpha)$.

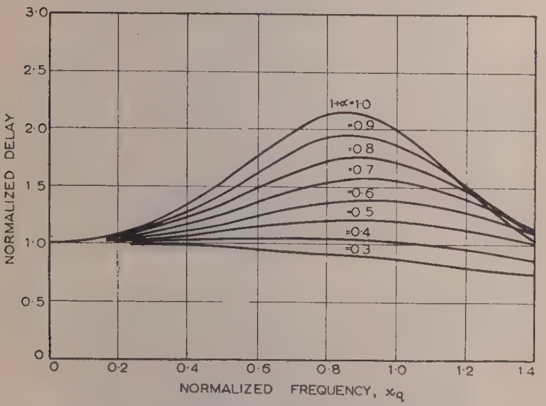


Fig. 10.—Normalized delay/frequency curve for bridged T-network for different values of $(1 + \alpha)$.

The gain characteristic, eqn. (11), shows a maximum within the pass band for values of $(1 + \alpha) > 0\cdot5$. For $(1 + \alpha) < 0\cdot5$, the characteristic falls steadily from the zero frequency value. A flat amplitude characteristic is obtained up to about 50% of the pass band (defined by the cut-off frequency ω_c) for $(1 + \alpha) = 0\cdot5$.

The delay characteristic, eqn. (13), is very similar to the gain characteristc. For values of $(1 + \alpha) > 0\cdot35$ it shows a peak within the pass band, and for $(1 + \alpha) < 0\cdot35$ the delay diminishes steadily with frequency from its zero-frequency value. If $(1 + \alpha) = 0\cdot35$ a very flat delay characteristic is obtained up to about 70% of the pass band.

A comparison of the curves in Figs. 9 and 10 shows that in the case of the bridged T-network very flat gain and delay characteristics cannot be obtained at the same time. If an amplifier is built to give a flat gain characteristic, its delay characteristic will not be very satisfactory. Also, since values of $(1 + \alpha)$ smaller than $0\cdot45$ require the coefficient of coupling between adjacent coils to be in excess of 38%, practical difficulties limit the construction of such sections of line. Thus it is not possible to use this network alone to obtain a very flat delay characteristic.

The bridged T-network may, however, be used very effectively to build an amplifier with a flat delay characteristic if the anode line is made up of m-derived sections and the grid line is built with bridged T-sections. (See Section 10.4 for the theoretical expressions for the gain of such amplifiers.) Thus, if it is adjusted so that the value of m for the anode line is $1\cdot4$, and $(1 + \alpha)$ for the grid line is $0\cdot45$, and also $R_{bg}C_{gk}/R_{0p}C_{pk} = 0\cdot6$ (R_{0p} and R_{bg} being the characteristic impedances, and C_{pk} and C_{gk} being the shunt capacitances of the anode and grid lines respectively), the resultant amplifier has a flat delay characteristic up to about 110% of the pass band (defined by cut-off frequency $\omega_c = 2/R_{0p}C_{pk}$), i.e. up to a frequency well beyond the -6 dB point. Also, the gain characteristic falls continuously from its zero-frequency value as the frequency increases. Such a characteristic is quite suitable from the point of view of the transient response of the amplifier. Calculated curves for the gain and phase shift of such an amplifier are shown in Fig. 11.

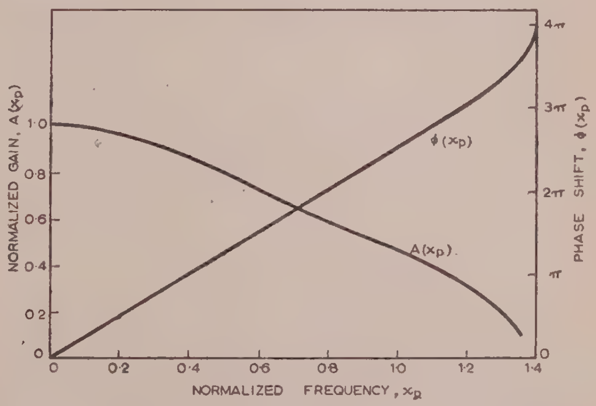


Fig. 11.—Normalized gain and phase-shift curves for amplifier with bridged T-network in the grid circuit and m-derived network in the anode circuit.

(6) EXPERIMENTAL RESULTS

In order to verify some of the calculations described above, distributed amplifiers of different models were built and their characteristics studied.

For the amplifier described in Section 3, having the constant-K LC network, a low-frequency model was built. The cut-off frequency of the amplifier was kept very low in order to avoid complications arising out of unwanted couplings between the inductance elements. For the same reason, closed-magnetic-circuit dust cores were used for the inductances of the lumped lines. In order to minimize the effects of stray capacitances, the shunt capacitances of the lines were made large compared with valve and wiring capacitances. These measures were taken in order to be able to observe the effects of staggering in the absence of major interfering effects. For the terminations of the anode and grid lines, m-derived half-sections were used. The characteristic impedance of the anode line was made 600 ohms. Three values of the staggering factor were taken ($q = 1\cdot0$, $0\cdot809$ and $0\cdot7$) and the grid line was successively adjusted to have the corresponding cut-off frequencies. The amplifier had 5 valves.

The steady-state gain and phase-shift characteristics obtained experimentally for the three values of the staggering factor are shown in Figs. 12 and 13.

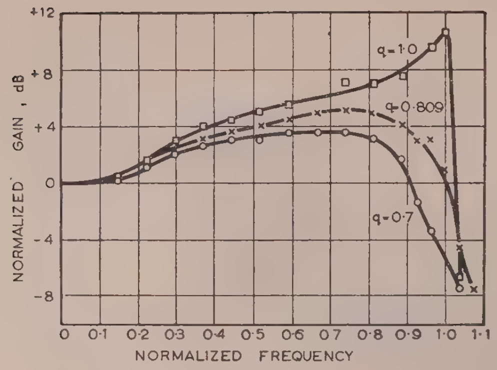


Fig. 12.—Experimental gain/frequency characteristic of 5-valve amplifier with constant-k LC filter network for different values of staggering factor q.

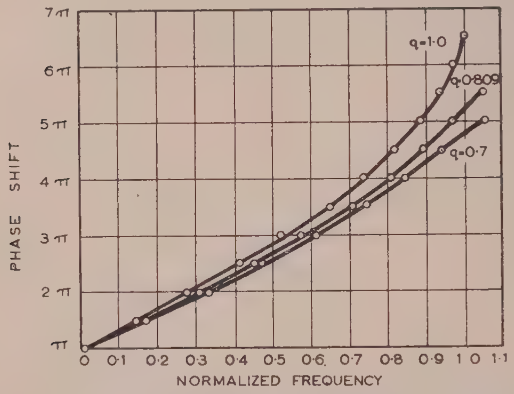


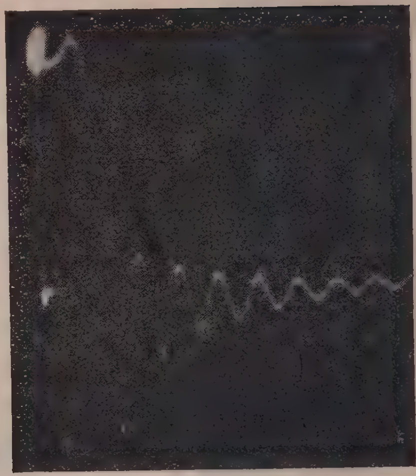
Fig. 13.—Experimental phase-shift/frequency characteristic of 5-valve amplifier with constant-k LC filter network for different values of staggering factor q.

Photographic records of the transient response of these amplifiers were also made. They are shown in Figs. 14(*a*), 14(*b*) and 14(*c*). The progressive reduction in the amount of overshoot and the decay time of the oscillations of the output voltage (for a step voltage input) with the reduction of the staggering factor is clearly noticeable. This is a consequence of the linearization of the phase-shift characteristic due to staggering. The rise time of the amplifier is also found to improve.

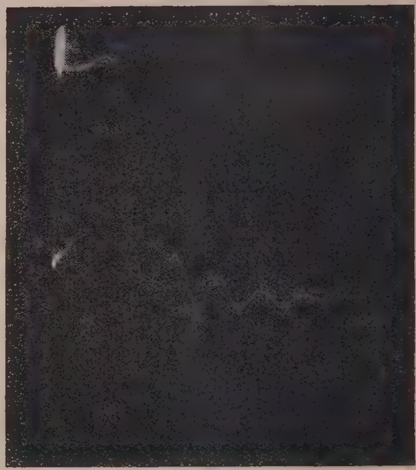
For the staggered amplifier described in Section 4, another low-frequency model was constructed. The anode-line cut-off frequency of this amplifier was 1·39 Mc/s, and air-core coils were used to obtain the required amount of coupling between the adjacent coils. In order to study the effects of staggering, the m-value of the anode line was kept low (1·14) and that of the grid line was adjusted successively to 1·14, 1·3 and 1·4. For such high values of m the required coupling between adjacent coils is very large. It may not be possible to attain these values if the coils are spaced uniformly, especially when the ratio of length to diameter of the individual coils is large. In the experimental amplifier the coupling was increased to the desired value by taking some turns of each coil very near to its neighbours. For the terminations of the anode and grid lines m-derived half-sections were used as in the previous case.

Photographic records of the transient response of the amplifier for these three combinations of the anode- and grid-line m-values were made. These are shown in Figs. 15(*a*), 15(*b*) and 15(*c*).

(*a*)



(*b*)



(*c*)

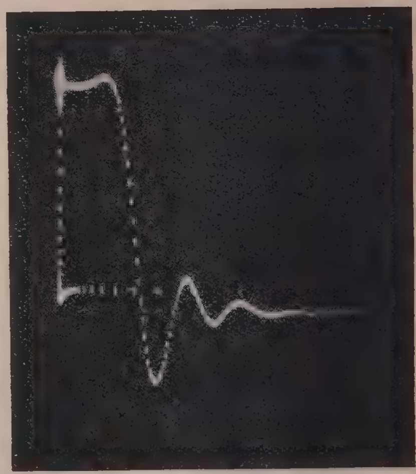
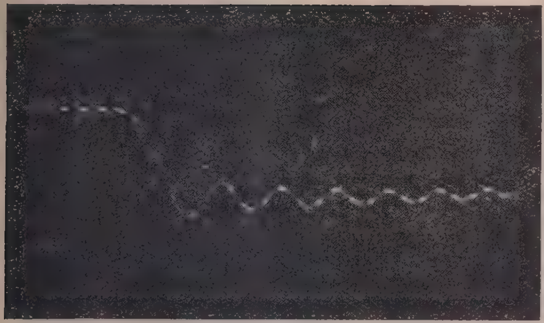


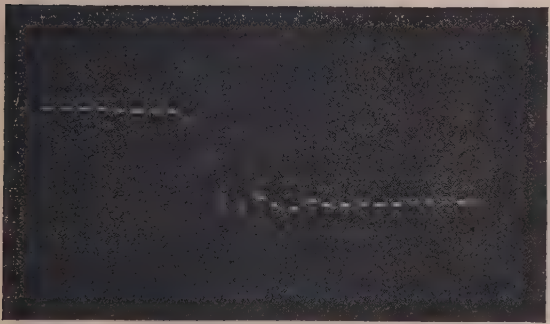
Fig. 14.—Photographic records of transient response obtained with step-voltage input on 5-valve amplifier with constant-k LC filter network.

(*a*) Staggering factor, $q = 1\cdot0$.
(*b*) Staggering factor, $q = 0\cdot809$.
(*c*) Staggering factor, $q = 0\cdot7$.



(*a*)

(*b*)



(*c*)

Fig. 15.—Photographic records of transient response obtained with a step-voltage input on 5-valve amplifier with *m*-derived filter sections for different combinations of anode- and grid-line *m*-values.

(*a*) $m_p = m_g = 1 \cdot 14$.
(*b*) $m_p = 1 \cdot 14$; $m_g = 1 \cdot 3$.
(*c*) $m_p = 1 \cdot 14$; $m_g = 1 \cdot 4$.

Time marks are 0·25 microsec apart.

As in the previous case, the overshoot, the decay time of the oscillations and the rise time show pronounced improvement as the stagger is increased.

The steady-state gain and phase-shift characteristics obtained by experiment corresponding to the last of the three combinations mentioned above (i.e. $m_p = 1 \cdot 14$ and $m_g = 1 \cdot 4$) are shown in Fig. 16, and the corresponding theoretical curves are plotted to the same scales to facilitate comparison.

The theoretical calculations for the steady-state characteristics are made on the assumption of lossless lines terminated with their proper characteristic impedances for all frequencies within

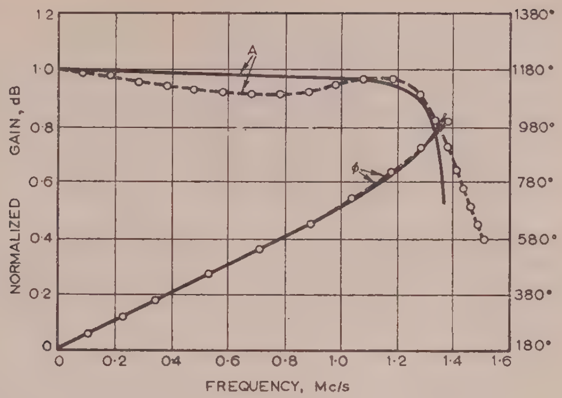


Fig. 16.—Gain and phase-shift characteristics for values of $m_p = 1 \cdot 14$ $m_g = 1 \cdot 4$.

——— Theoretical curves.
--o-- Experimental observations.

the pass band. As this assumption is not rigorously true, the experimental characteristics show some deviations from the calculated values.

(7) CONCLUSION

Phase staggering is found to be a very simple and useful device for improving the overall gain and phase-shift characteristics of different types of distributed amplifiers. The greatest improvement is obtained in the case in which *m*-derived sections form the anode and the grid lines.

(8) ACKNOWLEDGMENTS

The author is indebted to Prof. S. K. Mitra for his interest in the work, and to Mr. A. K. Chowdhury for constant help and guidance. The author's thanks are also due to Mr. N. B. Chakraborty for some very helpful discussions and to the Government of India for financial help.

(9) REFERENCES

(1) GINZTON, E. L., HEWLETT, W. R., JASBERG, J. H., and NOE, J. D.: "Distributed Amplification," *Proceedings of the Institute of Radio Engineers*, 1948, **36**, p. 956.

(2) BASSETT, H. G., and KELLY, L. C.: "Distributed Amplifiers: Some New Methods for Controlling Gain/Frequency and Transient Responses of Amplifiers having Moderate Bandwidths," *Proceedings I.E.E.*, Paper No. 1571 R, January, 1954 (**101**, Part III, p. 5).

(3) TISCHER, F.: "The Effective Bandwidth of Video Amplifiers," *Transactions of the Royal Institute of Technology, Stockholm*, 1953, No. 63.

(10) APPENDICES

(10.1) General Case of Staggered Distributed Amplifier

Consider the portion of a distributed amplifier between AB and CD in Fig. 17. It is the portion between the middle of a series element immediately preceding the grid of a valve and the middle of the series element immediately following the anode of the same valve. Let the voltage at CD due to a voltage V_i at AB be V, considering only the forward travelling signal. Let $V/V_i = G$, the gain of the system between AB and CD. Let there be n valves in all in the amplifier, and let the phase shifts per section of the anode line and grid line be θ_p and θ_g respectively, each section being a T-structure as in Fig. 17 between AB and QR.

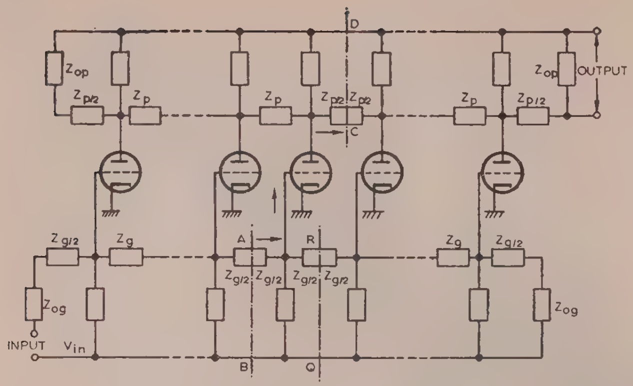


Fig. 17.—Simplified circuit diagram for calculating the effects of staggering.

Assuming that there is no loss in amplitude due to the lines within the pass band (but only phase shifts of θ_p and θ_g per section as mentioned above) the component signal V_1 at the output terminals (due to the input V_{in}) passing through the first valve of the amplifier and all but one section of the anode line may be written

$$V_1 = GV_{in}\varepsilon^{-j(n-1)\theta_p} \quad . \quad . \quad . \quad . \quad . \quad (14)$$

In a similar manner, the component signal V_2 passing through the second valve to the output terminals is

$$V_2 = GV_{in}\varepsilon^{-j[(n-2)\theta_p+\theta_g]} \quad . \quad . \quad . \quad . \quad (15)$$

since this component passes through one section of the grid line and $(n-2)$ sections of the anode line. The other components, passing through the third, fourth . . . nth valves are given by

$$V_3 = GV_{in}\varepsilon^{-j[(n-3)\theta_p+2\theta_g]}$$
$$V_4 = GV_{in}\varepsilon^{-j[(n-4)\theta_p+3\theta_g]}$$
$$\cdots$$
$$V_n = GV_{in}\varepsilon^{-j(n-1)\theta_g}$$

The resultant output V_0 is

$$V_0 = \sum_1^n V_n = GV_{in}\varepsilon^{-j(n-1)\theta_p}[1 + \varepsilon^{j\psi} + \varepsilon^{2j\psi} + \ldots + \varepsilon^{j(n-1)\psi}] \quad . \quad . \quad . \quad . \quad . \quad (16)$$

where ψ is written for $(\theta_p - \theta_g)$.

Summing,

$$V_0 = GV_{in}\varepsilon^{-j(n-1)\theta_p}\frac{(1-\varepsilon^{jn\psi})}{(1-\varepsilon^{j\psi})} = GV_{in}\frac{\sin n\psi/2}{\sin\psi/2}\varepsilon^{-j(n-1)\phi'} \quad . \quad (17)$$

where ϕ' is written for $(\theta_p + \theta_g)/2$. Since the angle of G proves to be equal to $-\phi'$ for all the types of network mentioned in the text, the overall phase shift is equal to $-n\phi'$, and is thus equal to the average of the phase shifts introduced by the anode and the grid lines. As a result, the overall delay characteristic is equal to the average of the delay characteristics of the anode line and grid line. The magnitude of G depends on the characteristic impedance and transfer characteristics of the network used.

The expressions for the gain characteristics of the different staggered amplifiers will now be deduced.

(10.2) Amplifier with Constant-*k* *LC* Filter Network

For an amplifier with constant-*k* *LC* filter network, the variation in the magnitude of G is the same as that of the mid-shunt characteristic impedance of a constant-*k* *LC* section. Thus

$$G = \frac{g_m R_0}{2}\frac{1}{\sqrt{(1-x^2)}}\varepsilon^{-j(\theta_p+\theta_g)/2} = \frac{g_m R_{0\pi}}{2}\varepsilon^{-j\phi'} \quad . \quad (18)$$

where $R_0 = \sqrt{(L/C)}$ for the anode line.

$x = \frac{\omega}{2}R_0 C_{pk} = \omega/\omega_{0p}$.

$\theta_p = 2 \arcsin x$; $\theta_g = 2 \arcsin qx$.

$q = \omega_{0p}/\omega_{0g}$.

g_m = transconductance of the valves used.

ω_{0p} and ω_{0g} = anode- and grid-line cut-off frequencies, respectively.

Applying the general relation, eqn. (17), the overall gain characteristics is

$$\frac{V_0}{V_{in}} = g_m\frac{Z_{0\pi}}{2}\frac{\sin n\psi/2}{\sin\psi/2}\varepsilon^{-jn\phi'}$$

where $\psi = (\theta_p - \theta_g)$

To calculate the gain from the grid of the first valve to the anode of the final valve, the half-T sections at the input and output ends are omitted and the expression becomes

$$\frac{V_0}{V_{in}} = g_m\frac{Z_{0\pi}}{2}\frac{\sin n\psi/2}{\sin\psi/2}\varepsilon^{-j(n-1)\phi'} \quad . \quad . \quad . \quad . \quad (19)$$

In order that the amplitude may fall to zero just below the cut-off frequency of the anode line, the factor

$$\frac{\sin n\psi/2}{\sin\psi/2}$$

has to vanish at ω_{0p}, that is,

$$(n\psi/2) = \pi \text{ at } x = 1 \quad . \quad . \quad . \quad . \quad . \quad (20)$$

Since $\psi/2 = \arcsin x - \arcsin Vx$, putting $x = 1$ and solving for q, we obtain from eqn. (20)

$$q = \sin(\pi/2 - \pi/n)$$

(10.3) Amplifier with *m*-derived *LC* Filter Network

For the determination of G, consider the circuit arrangement shown in Fig. 18. This is a portion of the equivalent circuit of

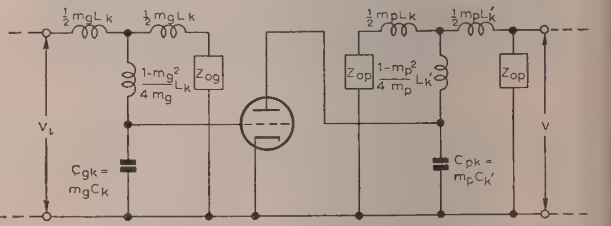


Fig. 18.—Equivalent circuit for calculating V/V_i for *m*-derived sections.

the *m*-derived *LC* filter used in the staggered distributed amplifier. The ratio V/V_i may be calculated by the method used by Ginzton *et al.*[1] This ratio becomes

$$G = \frac{V}{V_i} = \frac{R_{0p}g_m}{2}$$
$$\frac{1}{\{(1-x_g^2)[1-(1-m_g^2)x_g^2][1-(1-m_p^2)x_p^2]\}^{\frac{1}{2}}}\varepsilon^{-j(\theta_p+\theta_g)/2}$$
$$= g_m\frac{Z_0'}{2}\varepsilon^{-j\phi'} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

where $\theta_g = 2 \text{ arc tan } \dfrac{m_g x_g}{\sqrt{(1 - x_g^2)}}$.

$\theta_p = 2 \text{ arc tan } \dfrac{m_p x_p}{\sqrt{(1 - x_p^2)}}$.

(θ_g and θ_p are the phase shifts of the anode and grid lines respectively, per section.)

$x_g = \omega R_{0g} C_{gk}/2m_g$.

$x_p = \omega R_{0p} C_{pk}/2m_p$.

m_g and m_p = m-values of the grid and anode line, respectively.

R_{0p} = characteristic impedance of the anode line (at $\omega = 0) = \sqrt{(L'_k/C'_k)}$.

R_{0g} = characteristic impedance of the grid line (at $\omega = 0) = \sqrt{(L_k/C_k)}$.

Hence, the overall gain characteristic of an amplifier using this network and having n sections may be written from eqn. (17)

$$\frac{V_0}{V_{in}} = \frac{g_m Z'_0}{2} \frac{\sin n\psi/2}{\sin \psi/2} \varepsilon^{-jn\phi'} \quad . \quad . \quad . \quad . \quad (22)$$

where $\psi = (\theta_p - \theta_g)$

While combinations of the anode- and grid-line m-values are chosen from Fig. 6 in order to obtain a flat delay characteristic, it is assumed that the curves of these delay characteristics are normalized with respect to the same frequency. This requires that $x_p m_p$ be equal to $x_g m_g$ for a given value of ω, i.e. $R_{0g} C_{gk} = R_{0p} C_{pk}$. Thus let

$$x_g m_g = x_p m_p = x_q$$

then Z'_0 becomes

$$Z'_0 = \frac{R_{0p} m_g}{\{(m_g^2 - x_q^2)[(1 + x_q^2(1 - 1/m_g^2)][1 + x_q^2(1 - 1/m_p^2)]\}^{\frac{1}{2}}} \quad . \quad (23)$$

The corresponding gain characteristics of the unstaggered amplifiers are obtained by putting $m_g = m_p$ in all the above expressions, and eqns. (8) and (9) of Section 4 are the result.

(10.4) Amplifier with Bridged-T All-pass Sections in Grid Line and *m*-derived Low-pass Sections in Anode Line

For this case, the value of V/V_i, calculated in a way similar to those in the previous cases (see Fig. 19), is

$$\frac{V}{V_i} = \frac{g_m R_{0p}}{2}$$

$$\frac{1}{\sqrt{\{[1 - x_g^2(1 + \alpha)]^2 + x_{kg}^2\}}\sqrt{[1 - (1 - m^2)x_m^2]}} \varepsilon^{-j(\theta_p + \theta_g)/2}$$

$$= \frac{g_m Z''_0}{2} \varepsilon^{-j\phi''} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

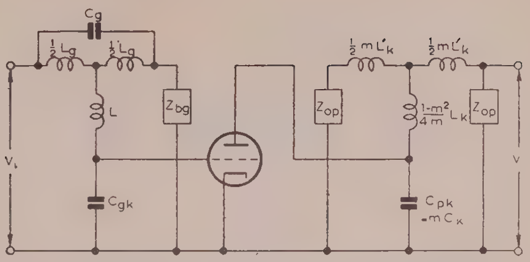


Fig. 19.—Equivalent circuit for calculating V/V_i for amplifier with bridged T-network in the grid circuit and m-derived sections in the anode circuit.

where $\theta_p = 2 \text{ arc tan } \dfrac{m x_m}{\sqrt{(1 - x_m^2)}}$

$\theta_g = 2 \text{ arc tan } \dfrac{x_{kg}}{1 - x_{kg}^2(1 + \alpha)}$

$x_{kg} = \dfrac{\omega}{2} R_{lg} C_{gk}$

$x_m = \dfrac{\omega}{2m} R_{0p} C_{pk}$

R_{0g}, R_{lg} = characteristic impedances of the anode and grid lines respectively, at zero frequency.

Applying the general relation [eqn. (17)] again, the overall gain characteristic is

$$\frac{V_0}{V_{in}} = \frac{g_m Z''_0}{2} \frac{\sin n\psi/2}{\sin \psi/2} \varepsilon^{-jn\phi''} \quad . \quad . \quad . \quad . \quad (25)$$

where $\psi = \theta_p - \theta_g$

Putting $m x_m = x_{kp}$,

$$Z''_0 = \frac{R_{0p}}{2\sqrt{\{[1 - x_{kg}^2(1 + \alpha)]^2 + x_{kg}^2\}}\sqrt{[1 + (1 - 1/m^2)x_{kp}^2]}} \quad . \quad (26)$$

so that the overall gain characteristic may be written

$$A(x_{kp}) = g_m \frac{Z''_0}{2} \frac{\sin n\psi/2}{\sin \psi/2} \quad . \quad . \quad . \quad . \quad (27)$$

and the overall phase-shift characteristic

$$\phi(x_{kp}) = \frac{n(\theta_p + \theta_g)}{2} \quad . \quad . \quad . \quad . \quad . \quad (28)$$

For the amplifier described in the paper, $m_p = 1 \cdot 4$, $(1 + \alpha) = 0 \cdot 45$, $(R_{lg} C_{gk}/R_{0p} C_{pk}) = 0 \cdot 605$.

621.372.8 : 621.3.011.21

Paper No. 1841 R
Sept. 1955

THEORY OF IMPERFECT WAVEGUIDES: THE EFFECT OF WALL IMPEDANCE

By A. E. KARBOWIAK, Ph.D.

(*The paper was first received* 27*th November*, 1954, *and in revised form* 29*th January*, 1955.)

SUMMARY

Wave propagation in waveguides whose walls exhibit an arbitrary but small surface impedance is studied.

It is shown that, although in the case of circular waveguides all E- and H-modes are stable, in the case of rectangular waveguides all E-modes and higher-order H-modes are unstable, leaving only H_0-modes that can propagate without change of form.

General formulae are derived from which the propagation coefficients of imperfect waveguides can be calculated, and for imperfect cavities an expression for the bandwidth and resonant frequency is obtained.

LIST OF PRINCIPAL SYMBOLS

μ_0, ϵ_0 = Permeability and permittivity of free space.

$k_0 = \dfrac{2\pi}{\lambda_0} = \omega\sqrt{(\mu_0\epsilon_0)}$ = Free-space propagation coefficient.

λ_0 = Free-space wavelength.

$\mu_i, \epsilon_i, \sigma_i$ = Permeability, permittivity and conductivity of the ith medium.

$k_i = \omega\sqrt{(\mu_i\epsilon_i)}$ = Propagation coefficient of the ith medium.

$Z_0 = \sqrt{\left(\dfrac{\mu_0}{\epsilon_0}\right)}$ = Free-space impedance.

$Z_i = \dfrac{1}{Z_0}\sqrt{\left(\dfrac{\mu_i}{\epsilon_i}\right)}$ = Normalized intrinsic impedance of the ith medium.

$Z_s = \left.\dfrac{E_t}{H_t}\right|_S$ = Normalized (with respect to Z_0) surface impedance.

Z_η, Z_ζ = Principal "components" of Z_s (defined in Section 3.3).

$\gamma = \alpha + j\beta = k_2$ = Axial propagation coefficient of the guided wave.

α = Attenuation coefficient (axial).

$\beta = \dfrac{2\pi}{\lambda_g}$ = Phase-change coefficient (axial).

$h_0 = \dfrac{2\pi}{\lambda_c}$ = Cut-off coefficient.

k_x, k_y = Propagation coefficient (Cartesian) in the transverse plane of a rectangular waveguide.

a, b = Dimensions of a rectangular guide.

u, v = General co-ordinates in the transverse plane of the waveguide.

s = Radius of a circular waveguide.

L = Length (axial) of a resonator.

τ = Wave coupling coefficient normalized with respect to Z_0.

$\Delta_1\Delta_2$ = Quantities defined by eqn. (86).

e_1, e_2, e_3 = Metrical coefficients.

In the modal notation: $M_{m,n,l}$ the indices are:

m = "Circumferential" index (0, 1, 2, . . .).

n = "Radial" index (×, ●, 0, 1, 2, . . .).

l = "Axial" index (×, ●, 0, 1, 2, . . .).

These denote that either the wave is evanescent (×) in that direction, or it is propagating (●) in that direction or that the wave exhibits a number (0, 1, 2, . . .) of anti-nodes within the particular dimension.

The rationalized M.K.S. system of units is used throughout and the sinusoidal time-factor $\varepsilon^{j\omega t}$ is implied. Further, excepting Section 6, all field quantities are understood to contain the factor $\varepsilon^{-\gamma z}$.

The quantity Z_0 is absorbed in the symbol H (magnetic-field vector) and consequently all impedances and coupling coefficients are normalized with respect to that quantity. As a result of this normalization the quantities involved in the formulae are dimensionless and the wavelength, for example, is measured in units in which one chooses to measure distances.

(1) INTRODUCTION

The concept of impedance between two points of an electric circuit is an old one, and Ohm's law in its original form provides us with a definition of impedance (Z) in terms of two measurable quantities—the current (I) and the potential difference (V). Although impedance is not a directly measurable quantity (being measured by the ratio V/I) it is an important quantity since in a way it describes quantitatively an electrical property of a body or a network between two chosen points, and as such it can be regarded as a physical characteristic of the body or the network considered. Undoubtedly, it is largely due to the last-mentioned feature that the impedance concept has been so successfully employed in many branches of engineering and applied sciences.

There has been, however, some delay and even reluctance, in the past, to extend the impedance concept to "field" problems (as distinct from "circuit" problems) of electrical engineering. One of the difficulties of adopting the impedance concept to field problems is the absence of conduction current, and even when the electromagnetic field is bounded (as it usually is), the current flowing in the bounding surface is not related uniquely to the field at a point but is a function of the total field configuration. Consequently no satisfactory definition in terms of voltage and current is possible, but the difficulty is not real, and a satisfactory definition of impedance (due to Schelkunoff[1]) in terms of the field vectors $\boldsymbol{E}$ and $\boldsymbol{H}$ can be constructed. Accordingly we define the impedance in the direction n at a point in an electromagnetic field by

$$Z = \frac{1}{2}\frac{\boldsymbol{E}^2}{\boldsymbol{S}.\boldsymbol{n}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where $\boldsymbol{S}$ is the complex Poynting vector.* The impedance so defined is termed "wave impedance" and the concept of wave impedance has been used by a number of investigators[2,3] in the solution of a variety of problems.

Here the concept of impedance is applied to the solution of boundary-value problems. In particular, wave propagation in waveguides whose walls are other than perfectly conducting is investigated in detail, with particular reference to those whose

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Dr. Karbowiak is with Standard Telecommunication Laboratories, Ltd.

* This is analogous to $Z = \dfrac{1}{2}\dfrac{V^2}{P}$ in electric-circuit theory, where P is the complex power flow.

geometry is either perfectly rectangular or perfectly circular. Throughout the paper, where quantitative formulae are involved, these are subject to one assumption only—namely that the waveguide surface impedance is small. Under this condition the problem for any physical waveguide can be solved in two separate steps. First, a solution is found for an arbitrary value of surface impedance for the given guide shape (Sections 4, 5 and 6). Secondly, an expression for the surface impedance of the guide surface is found in terms of the physical properties of the surface (Section 3) and this value of surface impedance is used, in the general formulae derived, to find the value of propagation coefficients.

(2) OUTLINE OF APPROACH

(2.1) General

The simplest boundary-value problem that one can encounter is that involving surfaces (S) on which the wave function (ψ) satisfies either of the two requirements given by the following equations:

$$[\psi]_{S_1} = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$\left[\frac{\partial}{\partial n}(\psi)\right]_{S_2} = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

In the case of eqn. (2) the surface S_1 has the property that the tangential component of the $\boldsymbol{E}$-vector (E_t) on the surface S_1 is zero, while with eqn. (3) the tangential component of the $\boldsymbol{H}$-vector (H_t) is made to vanish on the bounding surface S_2. Because of this property the surface S_1 can be regarded as a short-circuit surface, while the surface S_2 forms a virtual open-circuit to the field considered. Such boundary conditions are called perfect.

No physically realizable bounding surface can satisfy the conditions of eqns. (2) or (3), but many can be made to approach very closely to one of them, and this is the state of boundary that will occupy our attention. Primarily we shall be concerned with boundaries on which E_t, and thus the ratio E_t/H_t, is small,

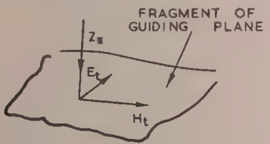


Fig. 1.—Sign convention for Z_S.

and this ratio* defines the surface impedance of the boundary (Fig. 1); it will be assumed throughout that the inequality (Section 3.1)

$$\left|\frac{E_t}{H_t}\right|_S = |Z_s| \ll 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

is satisfied.

The problem on hand is the wave propagation in imperfect uniform guides (rectangular, circular, etc.), and although the problem could, at least in principle, be solved for each particular case by rigorous solution of the boundary-value problem, the approach, in general, suffers from prohibitive complexity.

The alternative approach to the problem is well known in the case of metallic guides and provides us with a means of calculating the attenuation of the guide. In this method the problem is solved for the case of a perfect guide and *it is assumed that the field is not materially affected* by the substitution of imperfectly conducting walls for the perfectly conducting boundaries. Subsequently, the power loss to the guide wall is evaluated by an integration process and the integration of the Poynting vector over the cross-section of the guide yields the power carried by the guide; the attenuation of the guide is then calculated in the conventional manner.* This conventional approach to imperfect waveguides leads to an expression for the attenuation coefficient which can be correct only if the above italicized assumption is true.

The method introduced here is no more involved than the integration method and is based on the calculus of perturbation as explained below.

(2.2) Perturbation Effects

Suppose the solution to the problem of wave propagation in a perfect guide of a given shape is known and suppose that it is a mode of the region appropriate to the guide shape. Let this mode be denoted by M_0 and let $\beta_0\ (= -j\gamma_0)$ be the phase propagation coefficient of the guide; then the free-space propagation coefficient, k_0, and the separation constant, h_0 (cut-off coefficient), are connected as follows:

$$k_0^2 = h_0^2 + \beta_0^2 = h_0^2 - \gamma_0^2 \quad . \quad . \quad . \quad . \quad (5)$$

However, when the walls are imperfect the field inside the waveguide is given by

$$\psi = M_0 + \sum_{i=1}^{\infty} \tau_i M_i \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where M_i are orthogonal modes of the region. Further, h_0 will become h [equal to $h_0 + \delta(h)$] and consequently γ_0 will become γ (equal to $\gamma_0 + \delta\gamma$), and since k_0 is a constant,

$$\delta(\gamma) = \frac{h_0}{j\beta_0}\delta(h) \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

The perturbation term, $\delta(\gamma)$, is, in general, complex [$\delta(\gamma) = \alpha + j\delta(\beta)$], its real part α being the attenuation coefficient of the guide, and its imaginary part, $\delta(\beta)$, a measure of change of the guide wavelength ($\lambda_g = 2\pi/\beta$). The quantity $\delta(h)$ depends on Z_s, but the actual functional dependence is a matter of the guide geometry and can be calculated from the value of Z_s and the dimensions of the guide (see Sections 4 and 5).

(2.3) Purity and Stability of a Mode

With a perfect guide a simple function M_i is the solution to the problem—this is a pure mode. We note, however, that an imperfect guide cannot support a pure mode, but the original mode is, of necessity, accompanied by a denumerable infinity of other modes, as given by eqn. (6), and the field inside the guide is then said to be an impure M_0-mode.

It is thus evident that in the case of imperfect guides it is difficult to speak of modal propagation and the conventional division of waves into E- and H-modes takes less significance when viewed in this light. Yet, provided that as Z_s tends to zero all coefficients τ_i in eqn. (6) tend to zero, the mode M_0 is said to be stable and we can then speak of an imperfect guide as supporting a substantially pure M_0-mode.

(3) THE MEANING OF SURFACE IMPEDANCE

(3.1) Boundary separating Two Media

Imagine a plane TEM wave incident, at an angle θ to the normal, on a plane boundary between two homogeneous and isotropic media. As is well known, a proportion of the energy will be reflected and the remainder transmitted into the other medium (characteristic impedance, Z_2). Although the nature of

* This ratio will, in general, be complex, and carries a sign that is positive when the vector product ($\boldsymbol{E} \times \boldsymbol{H}$) of the field components at the surface is directed into the surface, and is negative when it is directed away from the surface.

* E.g. see LAMONT, R. L.: "Wave Guides" (Methuen and Co., Ltd., London).

the phenomenon will depend, in general, on the polarization of the incident wave, provided that $|Z_2| \ll |Z_1|$ the impedance "across" the plane separating the two media is substantially equal to Z_2, the characteristic impedance of medium 2, and, as such, is a constant characteristic (at any given frequency) of the boundary.

Now suppose medium 1 (a dielectric) filling the inner space of the guide is bounded by a metal crust whose thickness is large in comparison with the skin depth at the frequency considered: the solution to the wave propagation in such a guide is obtained by analysing wave propagation in a guide whose walls exhibit the physical property $Z_s = Z_2$, where Z_2 is the characteristic impedance of the metal concerned; at the same time the condition of eqn. (4) is satisfied.

Using the well-known expression for the characteristic impedance of a good conductor[4] we find that the surface impedance of a metal surface is given by

$$Z_s = R_s + jX_s = \frac{1}{Z_0}\sqrt{\left(\frac{\pi f \mu}{\sigma}\right)}(1 + j) = \frac{1}{Z_0}\sqrt{\left(\frac{\omega\mu}{\sigma}\right)}\angle 45^\circ \quad . \quad (8)$$

It will be observed that the surface impedance of a metal surface possesses equal resistive and inductive components. The resistive part is, in general, responsible for the attenuation while the reactive part is the factor influencing the phase velocity of the guided wave. As a result we can form the following physical picture of wave propagation in an imperfect guide.

The guide can be imagined to be formed by lossy walls possessing the attribute of resilience or inertia: the energy of the wave is thus slowly damped out and part of the energy carried by the wave is made to surge backwards and forwards across the guide wall. Further, because Z_s is finite, the energy of the wave is forced to penetrate the guide wall where, through the medium of the wall impedance, the principal mode is coupled to a number of other modes [eqn. (6)]: the purity of the mode is thus spoiled.

(3.2) The Uniform Impedance Sheet

A plane boundary separating two media as described above is an example of a homogeneous impedance sheet. In particular, since the value of this impedance is independent of the polarization of the field, it is not only a homogeneous but also an isotropic impedance sheet. There is a whole variety of such surfaces, but here we shall discuss but one more case that is of significance, namely that of a metal surface coated with a uniform layer of dielectric.

Imagine a TEM wave incident on a perfectly conducting plane coated with a thin layer of dielectric, as indicated in Fig. 2. It is evident that the wave impedance in the direction normal to the plane $y = 0$ depends on the value of the angle θ, and, in

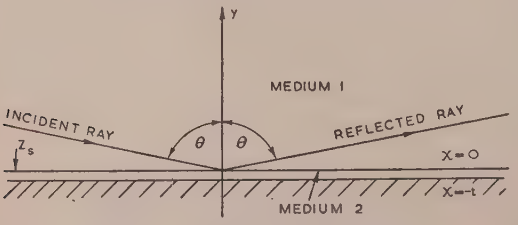


Fig. 2.—Dielectric coated surface.

general, no unique definition of the impedance of the plane $y = 0$ is possible. However, it can be shown that for large values of θ (glancing incidence) this impedance acquires a unique significance. This is of particular importance in connection with guided waves: here the electric vector is substantially normal to the surface ($\theta = 90°$).

Further, the surface impedance of a conducting sheet when coated with a thin layer of dielectric depends on the polarization of the wave. Thus, for example, this impedance when referred to the plane of the conductor is, for E-waves, given by (see Appendix)

$$Z_s = (k_0 t)\left[\left(\frac{h_1}{k_1}\right)^2 - \left(\frac{h_0}{k_0}\right)^2\right] \quad . \quad . \quad . \quad . \quad (9)$$

while for H-waves this impedance is zero. It must be emphasized, however, that this impedance is independent of the orientation of the impedance plane relative to the direction of wave propagation, and for this reason such an impedance sheet will be called isotropic.

(3.3) The Anisotropic Surface Impedance

Imagine a surface made up of strips of different materials as shown in Fig. 3, where the directions η and ζ are two principal

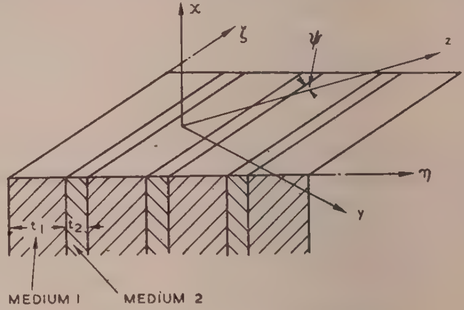


Fig. 3.—Anisotropic surface.

axes of the surface which need not, in general, coincide with the direction of the wave propagation z. It is clear that such a surface is no longer homogeneous; however, if the widths of the individual strips t_1 and t_2 are very small in comparison with the wavelength of the wave, the heterogeneous nature of this surface will be of no consequence (quasi-stationary solution) and the surface then may be expected to behave as if it were homogeneous, of a uniform impedance Z_s. The assumption that the period of the heterogeneity of the surface $(t_1 + t_2)$ is small in comparison with the wavelength will be made throughout this Section.

It is apparent from the impedance sheet, shown in Fig. 3, that the performance of such a structure will depend on the angle ψ, i.e. the orientation of the sheet relative to the direction of wave propagation; for this reason such a surface will be termed anisotropic. Further, it is evident that a single complex number will no longer suffice to describe the physical property (Z_s) of the surface adequately, and consequently the following notation will be introduced:

Let the impedance of the surface shown in Fig. 3 when the field is polarized with the **E**-vector directed along the ζ axis be $Z_s(\bar{\zeta}, \eta)$, and when polarized with the **E**-vector along η axis by $Z_s(\zeta, \bar{\eta})$; further, where no confusion is likely to arise these impedance components will be denoted by Z_ζ and Z_η, respectively.

Suppose, as an example, that the guide is made of thin strips of two different metals whose intrinsic impedances are Z_1 and Z_2 respectively; for a wave polarized along the η axis we have

$$Z_\eta = \frac{Z_1 t_1 + Z_2 t_2}{t_1 + t_2} \quad . \quad . \quad . \quad . \quad . \quad (10)$$

Similarly,

$$Z_\zeta = \frac{Z_1 Z_2 (t_1 + t_2)}{Z_2 t_1 + Z_1 t_2} \quad . \quad . \quad . \quad . \quad . \quad (11)$$

Another example of an anisotropic surface is a grooved or corrugated surface (Fig. 4). If we let Z_1 be the impedance (TEM; fringe field neglected) as presented by the narrow

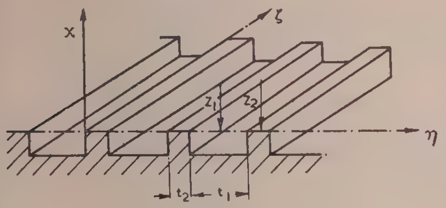


Fig. 4.—Corrugated surface.

groove, and Z_2 the intrinsic impedance of the metal of which the surface is made, and if the period of the structure is sufficiently small,

then
$$Z_\eta = \frac{Z_1 t_1 + Z_2 t_2}{t_1 + t_2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

Or, if the surface is made of a good conductor* ($|Z_2| \ll |Z_1|$),

$$Z_\eta \simeq Z_1 \frac{t_1}{t_1 + t_2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

while
$$Z_\zeta = Z_2 \frac{t_1 + t_2}{t_2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

(3.4) Other Types of Impedance Sheets

The impedance surfaces described above are just a few examples of surfaces that, although physically distinct, can be described in terms of the quantity Z_s, and in the case $|Z_s| \ll Z_0$ that quantity, characteristic of the surface, can be calculated quite simply as has been illustrated.

There is yet another type of surface that merits further elucidation: the randomly corrugated or "rough" surface. The reflection of electromagnetic waves from a rough surface has been discussed in detail by Rice[5] and Ament[6], and from their work some idea regarding the properties of rough guide surfaces may be obtained. The relevant properties are listed below:

(i) The roughness of a corrugated surface can be resolved into its Fourier spectrum; thus a periodic corrugated surface will have a discrete spectrum, while an aperiodic rough surface will be characterized by a continuous roughness spectrum.
(ii) If there is a significant correlation between a periodic function and the aperiodic rough surface, the roughness spectrum of the surface will exhibit a pronounced peak in the vicinity of that period, and, if this period is significantly less than the wavelength of the guided wave, such a surface tends to behave as if it were uniformly corrugated at the corresponding period.
(iii) The low-frequency components of the roughness spectrum, whose period is greater than the wavelength of the guided wave, are responsible for the scattering of the wave.
(iv) The high-frequency components of the roughness spectrum whose period is significantly less than the wavelength of the guided wave give rise to surface reactance.

(4) WAVE PROPAGATION IN RECTANGULAR WAVEGUIDES

(4.1) Field Equations pertaining to Rectangular Guides

(4.1.1) **E-waves.**

These are derived from the wave function

$$E_z = \left(\begin{matrix}\sin\\ \cos\end{matrix} k_x x\right)\left(\begin{matrix}\sin\\ \cos\end{matrix} k_y y\right) \quad . \quad . \quad . \quad . \quad (15)$$

* Compare this result with that obtained by a more rigorous method.[8]

so that the remaining field components, apart from a constant factor, are given by

$$\left.\begin{aligned}
H_x &= j\frac{k_0 k_y}{h^2}\left(\begin{matrix}\sin\\ \cos\end{matrix} k_x x\right)\left(\begin{matrix}\cos\\ -\sin\end{matrix} k_y y\right)\\
E_y &= j\frac{k_z k_y}{h^2}\left(\begin{matrix}\sin\\ \cos\end{matrix} k_x x\right)\left(\begin{matrix}-\cos\\ \sin\end{matrix} k_y y\right)\\
H_y &= j\frac{k_0 k_x}{h^2}\left(\begin{matrix}-\cos\\ \sin\end{matrix} k_x x\right)\left(\begin{matrix}\sin\\ \cos\end{matrix} k_y y\right)\\
E_x &= j\frac{k_z k_x}{h^2}\left(\begin{matrix}-\cos\\ \sin\end{matrix} k_x x\right)\left(\begin{matrix}\sin\\ \cos\end{matrix} k_y y\right)
\end{aligned}\right\} \quad . \quad . \quad (16)$$

(4.1.2) **H-waves.**

These are derived from the wave function

$$H_z = \left(\begin{matrix}\cos\\ \sin\end{matrix} k_x x\right)\left(\begin{matrix}\cos\\ \sin\end{matrix} k_y y\right) \quad . \quad . \quad . \quad . \quad (17)$$

so that the remaining field components, apart from a constant factor, are given by

$$\left.\begin{aligned}
H_x &= j\frac{k_z k_x}{h^2}\left(\begin{matrix}\sin\\ -\cos\end{matrix} k_x x\right)\left(\begin{matrix}\cos\\ \sin\end{matrix} k_y y\right)\\
E_y &= j\frac{k_0 k_x}{h^2}\left(\begin{matrix}-\sin\\ \cos\end{matrix} k_x x\right)\left(\begin{matrix}\cos\\ \sin\end{matrix} k_y y\right)\\
H_y &= j\frac{k_z k_y}{h^2}\left(\begin{matrix}\cos\\ \sin\end{matrix} k_x x\right)\left(\begin{matrix}\sin\\ -\cos\end{matrix} k_y y\right)\\
E_x &= j\frac{k_0 k_y}{h^2}\left(\begin{matrix}\cos\\ \sin\end{matrix} k_x x\right)\left(\begin{matrix}\sin\\ -\cos\end{matrix} k_y y\right)
\end{aligned}\right\} \quad . \quad . \quad . \quad (18)$$

In these expressions k_x, k_y and k_z have the following meaning (Fig. 5):

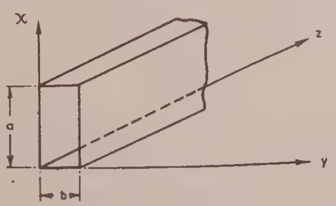


Fig. 5.—Co-ordinate system in relation to a rectangular guide.

For perfect guides

$$\left.\begin{aligned}
h_0^2 &= k_{x0}^2 + k_{y0}^2\\
k_0^2 &= k_{x0}^2 + k_{y0}^2 + k_{z0}^2\\
k_{x0} &= \left(\frac{m\pi}{a}\right),\ k_{y0} = \left(\frac{n\pi}{b}\right)
\end{aligned}\right\} \quad . \quad . \quad . \quad (19)$$

$$k_0 = \left(\frac{2\pi}{\lambda_0}\right),\ k_{z0} = \beta_0 = \frac{2\pi}{\lambda_g} \quad . \quad . \quad . \quad . \quad (20)$$

while for an imperfect guide these values are

$$\left.\begin{aligned}
k_0^2 &= k_x^2 + k_y^2 + k_z^2\\
k_x &= k_{x0} + \delta(k_x)\\
k_y &= k_{y0} + \delta(k_y)\\
k_z &= k_{z0} + \delta(k_z)\\
h &= h_0 + \delta(h)\\
\gamma &= \alpha + j\beta = [\alpha + j\delta(\beta)] + j\beta_0\\
\delta\gamma &= \alpha + j\delta\beta
\end{aligned}\right\} \quad . \quad . \quad (21)$$

The surface impedance of guide walls at $x = a$ and $y = b$ are

$$\left.\begin{aligned} \left.\frac{E_y}{H_z}\right|_{x=a} &= Z_s(\bar{y}, z) \\ -\left.\frac{E_z}{H_y}\right|_{x=a} &= Z_s(y, \bar{z}) \\ \left.\frac{E_z}{H_x}\right|_{y=b} &= Z_s(x, \bar{z}) \\ -\left.\frac{E_x}{H_z}\right|_{y=b} &= Z_s(\bar{x}, z) \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (22)$$

with analogous expressions for the surface impedance at $x = 0$ and $y = 0$.

(4.2) The E_{on} and H_{on}-waves

These waves are characterized by $k_{y0} = 0$. As is well known, a perfect guide cannot support an E_0-mode, thus the E_0-mode can be coupled to the H_0-mode only by the medium of, for example, the wall impedance. Consequently the field in a rectangular guide must be derived from

$$\left.\begin{aligned} H_z &= \cos k_x x \,.\, \cos k_y y \\ E_z &= \tau \sin k_x x \,.\, \sin k_y y \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (23)$$

where, since Z_s is small, τ and k_y ($= \delta k_y$) are likewise small.

We now suppose that all walls are perfect but for the wall at $x = a$, which will have two components: the circumferential component $Z_s(\bar{y}, z)$ and the axial component $Z_s(y, \bar{z})$. In this case obviously $k_y = \delta k_y = 0$ and there is no coupled E_0-wave; however, the H_0-wave, although its purity is unimpaired, suffers a change in the coefficients k_x and k_z as follows. From eqns. (22), (23) and (18) we have, to the first order of quantities,

$$Z_s(\bar{y}, z) = \left.\frac{E_y}{H_z}\right|_{x=a} = j\frac{k_0 k_x}{h_0^2}(-\sin k_x x)$$

$$\simeq -j\frac{k_0}{h_0}a\delta(h)$$

since $k_x = h$.

Consequently $$\delta(h) = \frac{1}{a}\left(\frac{h_0}{k_0}\right) jZ_s(\bar{y}, z) \quad . \quad . \quad . \quad . \quad (24)$$

Thus from eqn. (7),

$$\delta(\gamma)_1 = \frac{1}{a}\frac{h_0^2}{k_0\beta_0}[Z_s(\bar{y}, z)] \quad . \quad . \quad . \quad . \quad (25)$$

Therefore, the performance of the guide depends only upon the circumferential component of the impedance.

We turn our attention now to the wall at $y = b$, which again has two components: the circumferential $Z_s(\bar{x}, z)$ and the axial $Z_s(x, \bar{z})$. In this case $\delta k_x = 0$ while $k_y = \delta(k_y)$, and consequently there will be a coupling between the H_0- and E_0-waves. Neglecting terms of second and higher orders in $\delta(k_y)$, we get

$$Z_a = Z_s(x, \bar{z}) = \left.\frac{E_z}{H_x}\right|_{y=b} = \frac{\tau \,.\, b \,.\, \delta k_y}{j\frac{k_z k_x}{h_0^2} + \tau j\frac{k_0}{h_0^2}\delta k_y}$$

$$\simeq -j\tau\frac{h_0^2}{k_z k_x}b\delta k_y \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

and

$$-Z_c = -Z_s(\bar{x}, z) = \left.\frac{E_x}{H_z}\right|_{x=a} = \left(j\frac{k_0\delta k_y}{h_0^2} - j\tau\frac{k_z k_x}{h_0^2}\right)b\delta k_y \quad . \quad (27)$$

Substituting from eqns. (26) and (27) we get

$$b(\delta k_y)^2 = \left[jZ_c\left(\frac{h_0}{k_0}\right)^2 + jZ_a\left(\frac{\beta_0}{k_0}\right)^2\right]k \quad . \quad . \quad (28)$$

Since $$\delta(\gamma) = \frac{1}{2}\frac{\delta^2(k_y)}{j\beta_0} \quad . \quad . \quad . \quad . \quad . \quad (29)$$

therefore $$\delta(\gamma) = \frac{1}{2b}\left[Z_c\left(\frac{h_0}{k_0}\right)^2 + Z_a\left(\frac{\beta_0}{k_0}\right)^2\right]\frac{k_0}{\beta_0} \quad . \quad . \quad (30)$$

For a homogeneous surface impedance that is independent of the polarization, we have $Z_c = Z_a = Z_s$, for which case eqn. (30) becomes

$$\delta(\gamma)_2 = \frac{1}{2b}\frac{k_0}{\beta_0}(Z_s) \quad . \quad . \quad . \quad . \quad . \quad (31)$$

If in addition $Z_s(\bar{y}, z) = Z_s$, then adding eqns. (25) and (31) and multiplying by two, the total $\delta(\gamma)$ for a homogeneous rectangular guide is given by

$$\delta(\gamma)_T = \frac{k_0}{\beta_0}\left[\frac{2}{a}\left(\frac{h_0}{k_0}\right)^2 + \frac{1}{b}\right]Z_s \quad . \quad . \quad . \quad . \quad (32)$$

The real part of this expression is

$$\alpha = \frac{\lambda_g}{\lambda_0}\left[\frac{2}{a}\left(\frac{\lambda_0}{\lambda_c}\right)^2 + \frac{1}{b}\right]R_s \quad . \quad . \quad . \quad . \quad (33)$$

which is the familiar expression for the attenuation coefficient of a uniform metal guide.

An interesting property of the guide is that the coupling coefficient, τ, between the E_0- and H_0-waves depends solely on the axial component of the surface impedance of the broad face. From eqns. (26) and (28) we get

$$\tau = \frac{k_z}{h_0}\frac{jZ_a}{\left[jZ_c\left(\frac{h_0}{k_0}\right)^2 + jZ_a\left(\frac{\beta_0}{k_0}\right)^2\right]^{\frac{1}{2}}[bk_0]^{\frac{1}{2}}} \quad . \quad . \quad (34)$$

This, for the case $Z_c = Z_a = Z_s$, simplifies to

$$\tau = \frac{k_z}{h_0}\sqrt{\left(\frac{jZ_s}{bk_0}\right)} \quad . \quad . \quad . \quad . \quad . \quad (35)$$

which is a small quantity if Z_s is small.

Further, from eqn. (32) we conclude that while the attenuation coefficient, α, depends only on R_s, the phase propagation coefficient, β, is affected only [by the amount $\delta(\beta)$] by the reactive part of the surface impedance.

(4.3) Higher-Order Waves

Higher-order waves in rectangular guides are $H_{m,n}$- or $E_{m,n}$-waves (m, $n \neq 0$) for which $k_{y0} \neq 0$ and $k_{x0} \neq 0$ and the field of these waves is derived, when the walls at $x = 0$ and $y = 0$ are perfect, from eqns. (23). If we let τ be the normalized coupling coefficient then the field components are

$$\left.\begin{aligned} H_z &= \cos k_x x \cos k_y y \\ E_z &= \tau \sin k_x x \sin k_y y \\ E_y &= -j\left(\frac{k_0 k_x}{h^2} + \frac{k_z k_y}{h^2}\tau\right)\sin k_x x \cos k_y y \\ H_y &= j\left(\frac{k_z k_y}{h^2} - \frac{k_0 k_x}{h^2}\tau\right)\cos k_x x \sin k_y y \\ E_x &= j\left(\frac{k_0 k_y}{h^2} - \frac{k_z k_x}{h^2}\tau\right)\cos k_x x \sin k_y y \\ H_x &= j\left(\frac{k_z k_x}{h^2} + \frac{k_0 k_y}{h^2}\tau\right)\sin k_x x \cos k_y y \end{aligned}\right\} \quad . \quad (36)$$

Suppose now that all guide walls are perfect except the wall at $x = a$, which has a uniform surface impedance Z_1.

Thus
$$-\left.\frac{E_z}{H_y}\right|_{x=a} = Z_1 = \left.\frac{E_y}{H_z}\right|_{x=a} \quad . \quad . \quad . \quad . \quad (37)$$

To the order of $\delta(k_x)$ only, the substitution of eqn. (36) into eqn. (37) leads to

$$\left.\begin{aligned} \frac{\tau \,.\, a \,.\, \delta(k_x)}{j\left(\frac{k_z k_y}{h_0^2} - \frac{k_0 k_x}{h_0^2}\tau\right)} &= -Z_1 \\ j\left(\frac{k_0 k_x}{h_0^2} + \frac{k_z k_y}{h_0^2}\tau\right)a \,.\, \delta(k_x) &= Z_1 \end{aligned}\right\} \quad . \quad . \quad . \quad (38)$$

The elimination of Z_1 from eqn. (38) furnishes a quadratic for the coupling coefficient τ, with two solutions, namely

$$\left.\begin{aligned} \tau_1 &= \frac{k_0 \,.\, k_y}{k_z \,.\, k_x} \\ \tau_2 &= -\frac{k_z k_x}{k_0 k_y} = -\frac{1}{\tau_1} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (39)$$

It is important to note that, provided Z_1 is small, as has been assumed, τ_1 and τ_2 are independent of the value of Z_1; the coupling coefficient depends only on the order of the mode and the dimensions of the guide.

Substitution of eqn. (39) back into eqn. (38) gives

$$\left.\begin{aligned} \delta(k_x)_1 &= \frac{1}{a}\frac{k_x}{k_0}(jZ_1) \\ \delta(k_x)_2 &= \frac{1}{a}\frac{k_0}{k_x}(jZ_1) \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (40)$$

Consequently

$$\left.\begin{aligned} \delta(\gamma)_1 &= \frac{1}{a}\frac{k_x^2}{k_0\beta_0}Z_1 \\ \delta(\gamma)_2 &= \frac{1}{a}\frac{k_0}{\beta_0}Z_1 \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (41)$$

Since the coupling coefficient as given by eqn. (39) is not small we draw the conclusion that neither E- nor H-modes are stable with respect to a small perturbation of the quality of the boundary, unless m or n is zero.

Although no higher E- or H-mode can be stable in a rectangular guide, we see from eqns. (39) and (40) that the combination of waves

$$(E_x H)_{m,n} = (H_{m,n} + \tau_1 E_{m,n}) \quad . \quad . \quad . \quad . \quad (42)$$

or

$$(EH_x)_{m,n} = (H_{m,n} + \tau_2 E_{m,n}) \quad . \quad . \quad . \quad . \quad (43)$$

are stable. The propagation coefficients of these waves are respectively

$$\left.\begin{aligned} \gamma_1 &= \gamma_0 + \delta(\gamma)_1 \\ \gamma_2 &= \gamma_0 + \delta(\gamma)_2 \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (44)$$

Thus

$$\left.\begin{aligned} \beta_1 &= \beta_0 + \delta\beta_1 = \beta_0 + \frac{1}{a}\frac{k_x^2}{k_0\beta_0}X_1 \\ \beta_2 &= \beta_0 + \delta\beta_2 = \beta_0 + \frac{1}{a}\frac{k_0}{\beta_0}X_1 \end{aligned}\right\} \quad . \quad . \quad (45)$$

and

$$\left.\begin{aligned} \alpha_1 &= \frac{1}{a}\frac{k_x^2}{k_0\beta_0}R_1 \\ \alpha_2 &= \frac{1}{a}\frac{k_0}{\beta_0}R_1 \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (46)$$

Undoubtedly we can superimpose $(E_x H)_{m,n}$ and $(EH_x)_{m,n}$ in suitable proportions to get at any particular point along the guide a pure $E_{m,n}$- or $H_{m,n}$-wave, but since $(E_x H)$- and (EH_x)-modes have different phase velocities the wave will propagate down the guide suffering a change of form. By this we mean that, for example, an E-wave will, while passing down the guide, gradually change entirely into an H-wave and then back again into an E-wave. It is only the particular combination of E- and H-modes as given by eqns. (42) or (43) that is stable and can proceed down a rectangular guide without change of form. Further, we note that since $\alpha_1 = \alpha_2\left(\frac{k_x}{k}\right)^2$, and that since usually $a > \lambda_0/2$, α_1 is less than α_2 and consequently—unless the wave at the entrance to a guide is a pure (EH_x)-wave—the wave at the end of a long guide will be substantially a pure $(E_x H)$-mode.

The particular combination of E- and H-modes, as given by eqns. (42) and (43), with coupling coefficients as given by eqn. (39), are known as longitudinal-section waves, first mentioned by Buchholz[7] in 1939. These waves are characterized by the absence of the E_x component as in the $(E_x H)$-mode or the absence of the H_x component as in the (EH_x)-mode—as a study of eqn. (36) together with eqn. (39) will reveal. Fig. 6 illustrates the E and H lines of force of these modes in the transverse plane of the guide.

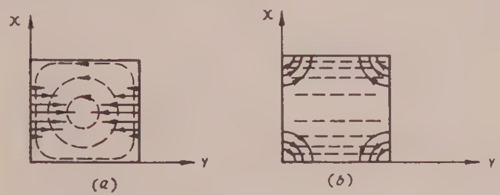


Fig. 6.—Field of longitudinal-section waves.

(a) $(E_x H)_{11}$-mode.
(b) $(EH_x)_{11}$-mode.
———— Electric lines of force.
— — — — Magnetic lines of force.

In conclusion we may state that it is meaningless to talk of attenuation of metal guides when carrying, for example, an E-mode. A metal surface has a surface impedance given by eqn. (8), and consequently it follows from the foregoing analysis that a metal guide cannot support an E-mode of whatever order. For a similar reason it is equally inappropriate to talk of attenuation of metal guides when supporting higher-order H-modes.

This study is not meant to be a challenge to the conventional method of evaluating the attenuation coefficient of a guide by the integration method as outlined in Section 2.1, but when the method is applied to modes other than H_0 a misleading set of results is obtained.

To complete our conclusions, consider a rectangular guide whose walls are perfect with the exception of the wall at $y = b$, which has a uniform surface impedance Z_2.

Thus
$$\left.\frac{E_z}{H_x}\right|_{y=b} = Z_2 = -\left.\frac{E_x}{H_z}\right|_{y=b} \quad . \quad . \quad . \quad . \quad (47)$$

Since the analysis for this case is similar to the one considered in eqn. (37) the details of the analysis will be omitted and only the results given. We find that, as before, there are two coupling coefficients, namely

$$\left.\begin{aligned} \tau_1' &= -\frac{k_0 k_x}{k_z k_y} \\ \text{and} \qquad \tau_2' &= \frac{k_z k_y}{k_0 k_x} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (48)$$

The corresponding perturbation terms are

$$\left.\begin{aligned}\delta(k_y)_1 &= \frac{1}{b}\left(\frac{k_y}{k_0}\right)(jZ_2)\\ \delta(k_y)_2 &= \frac{1}{b}\left(\frac{k_0}{k_y}\right)(jZ_2)\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (49)$$

and

$$\left.\begin{aligned}\delta(\gamma)_1 &= \frac{1}{b}\frac{k_y^2}{k_0\beta_0}Z_2\\ \delta(\gamma)_2 &= \frac{1}{b}\frac{k_0}{\beta_0}Z_2\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (50)$$

The only stable waves are

$$(E_yH)_{m,n} = (H_{m,n} + \tau_1' E_{m,n}) \quad . \quad . \quad . \quad . \quad (51)$$

and

$$(EH_y)_{m,n} = (H_{m,n} + \tau_2' E_{m,n}) \quad . \quad . \quad . \quad . \quad (52)$$

Finally, consider a guide whose walls at $x = 0$ and $y = 0$ are perfect, but the walls at $x = a$ and $y = b$ have impedances Z_1 and Z_2 respectively.

In this case we find not only that higher-order E- or H-modes cannot exist but that no linear combination of the modes (of the form $E_{m,n} + \tau H_{m,n}$) can be made to travel down a uniform rectangular guide—whose two adjacent walls are imperfect—without change of form. In fact it is only the H_{0n}- and H_{m0}-modes that can propagate down a uniform and imperfect guide without change of form.

(5) APPLICATION TO CIRCULAR GUIDES

(5.1) Field Equations pertaining to Circular Waveguides

(5.1.1) E-waves.

$$\left.\begin{aligned}E_z &= J_m(hr)\binom{\cos}{\sin}m\phi\\ E_r &= -\frac{j\beta}{h}J_m'(hr)\binom{\cos}{\sin}m\phi\\ E_\phi &= \frac{j\beta}{h^2}\frac{m}{r}J_m(hr)\binom{\sin}{-\cos}m\phi\\ H_r &= -\frac{jk_0}{h^2}\frac{m}{r}J_m(hr)\binom{\sin}{-\cos}m\phi\\ H_\phi &= -\frac{jk_0}{h}J_m'(hr)\binom{\cos}{\sin}m\phi\end{aligned}\right\} \quad . \quad . \quad . \quad (53)$$

(5.1.2) H-waves.

$$\left.\begin{aligned}H_z &= J_m(hr)\binom{\cos}{\sin}m\phi\\ E_r &= \frac{jk_0}{h^2}\frac{m}{r}J_m(hr)\binom{\sin}{-\cos}m\phi\\ E_\phi &= \frac{jk_0}{h}J_m'(hr)\binom{\cos}{\sin}m\phi\\ H_r &= -\frac{j\beta}{h}J_m'(hr)\binom{\cos}{\sin}m\phi\\ H_\phi &= \frac{j\beta}{h^2}\frac{m}{r}J_m(hr)\binom{\sin}{-\cos}m\phi\end{aligned}\right\} \quad . \quad . \quad . \quad (54)$$

In these expressions the coefficients h and k_0 have the following meanings (Fig. 7). For perfect guides:

$$k_0^2 = h_0^2 + \beta_0^2 = h_0^2 - \gamma_0^2 \quad . \quad . \quad . \quad . \quad (55)$$

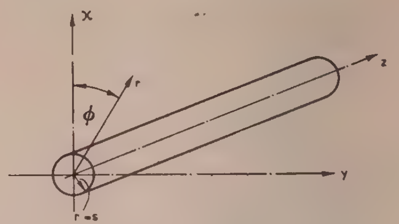


Fig. 7.—Cylindrical polar co-ordinate system.

where the coefficient h_0 is given for E-waves as the solution of $J_m(h_0 s) = 0$, and for H-waves as the solution of $J_m'(h_0 s) = 0$, and where s is the radius of the guide.

For an imperfect guide the coefficients are given by

$$\left.\begin{aligned}k_0^2 &= h^2 - \gamma^2\\ h &= h_0 + \delta h\\ \gamma &= \gamma_0 + \delta\gamma = j\beta_0 + \delta\gamma\\ \delta\gamma &= \alpha + j\delta\beta\end{aligned}\right\} \quad . \quad . \quad . \quad (56)$$

The surface impedance of the guide wall at $r = s$ is

$$\left.\begin{aligned}-\frac{E_z}{H_\phi}\bigg|_{r=s} &= Z_s(\phi, \bar{z}) = Z_z\\ &= \text{axial impedance}\\ \frac{E_\phi}{H_z}\bigg|_{r=s} &= Z_s(\bar{\phi}, z) = Z_\phi\\ &= \text{circumferential impedance}\end{aligned}\right\} \quad . \quad (57)$$

(5.2) E-waves

Let the field of an $E_{m,n}$-wave inside a perfect guide be derived from

$$E_z = J_m(hr)\cos(m\phi) \quad . \quad . \quad . \quad . \quad . \quad (58)$$

Then, in the case of an imperfect guide the wall impedance will couple, in general, a certain amount of H-wave, which is derived from

$$H_z = \tau J_m(hr)\sin(m\phi) \quad . \quad . \quad . \quad . \quad . \quad (59)$$

where τ is the normalized coupling coefficient and h is given by eqn. (56) with h_0 as the solution of

$$J_m(h_0 s) = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (60)$$

In this case, substituting from eqns. (53) and (54) into eqn. (57), we get

$$-Z_z = \frac{J_m(hs)}{-j\frac{k_0}{h_0}J_m'(hs) + \tau j\frac{\beta_0}{h_0^2}\frac{m}{r}J_m(hs)} \quad . \quad . \quad (61)$$

$$Z_\phi = \frac{j\frac{\beta_0}{h_0^2}\frac{m}{r}J_m(hs) + \tau j\frac{k_0}{h_0}J_m'(hs)}{\tau J_m(hs)} \quad . \quad . \quad . \quad (62)$$

We now expand in Taylor's series the Bessel functions occurring in eqns. (61) and (62) about the point $h_0 s$, and to the first order of quantities we get

$$\left.\begin{aligned}J_m(hs) &\simeq s\delta(h)J_m'(h_0 s)\\ \text{and} \qquad J_m'(hs) &\simeq J_m'(h_0 s)\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (63)$$

On substituting eqn. (63) into eqns. (61) and (62) and neglecting higher-order quantities, we get

$$\tau = \frac{\beta_0}{h_0^2}\frac{m}{s}(jZ_z) \quad . \quad . \quad . \quad . \quad . \quad . \quad (64)$$

and
$$\delta(h) = \frac{1}{s}\frac{k_0}{h_0}(jZ_z) \quad . \quad . \quad . \quad . \quad . \quad . \quad (65)$$

Thus
$$\delta(\gamma) = \frac{1}{s}\frac{k_0}{\beta_0}Z_z \quad . \quad . \quad . \quad . \quad . \quad . \quad (66)$$

Consequently, separating real and imaginary parts, the attenuation coefficient, α, is given by

$$\alpha = \frac{1}{s}\frac{\lambda_g}{\lambda_0}R_z \quad . \quad . \quad . \quad . \quad . \quad . \quad (67)$$

and the phase propagation coefficient, β_0, is altered by

$$\delta(\beta) = \frac{1}{s}\frac{\lambda_g}{\lambda_0}X_z \quad . \quad . \quad . \quad . \quad . \quad . \quad (68)$$

The coupling coefficient, τ, is small and tends to zero as $|Z_z| \to 0$, consequently all E-waves in a circular guide are stable. Further, since for all E_0-modes $\tau = 0$, we conclude that the purity of E_0-modes is unimpaired by the finite value of Z_z.

It is important to note that the performance of a circular guide is unaffected by the value of the circumferential impedance when small, and it is tempting to deduce that the propagation of E-waves is unaffected by the circumferential impedance, however large. An analysis carried out to that end shows that this is indeed the case.

(5.3) H-waves

Let the field inside a perfect guide, when carrying an $H_{m,n}$-mode, be derived from

$$H_z = J_m(hr)\cos(m\phi) \quad . \quad . \quad . \quad . \quad . \quad (69)$$

Then, if the guide is imperfect the wall impedance will couple, in general, a certain amount of E-wave, which is derived from

$$E_z = \tau J_m(hr)\sin(m\phi) \quad . \quad . \quad . \quad . \quad . \quad (70)$$

In these expressions τ is the normalized coupling coefficient and h is given by eqn. (56) with h_0 as the solution of

$$J'_m(h_0 s) = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (71)$$

Proceeding as before, from eqn. (57) on substitution of eqns. (53) and (54), we get

$$-Z_z = \frac{\tau J_m(hs)}{j\frac{\beta_0}{h_0}\frac{m}{r}J_m(hs) - \tau j\frac{k_0}{h_0}J'_m(hs)} \quad . \quad . \quad . \quad (72)$$

and
$$Z_\phi = j\frac{k_0}{h_0}\frac{J'_m(hs)}{J_m(hs)} - \tau j\frac{\beta_0}{h_0^2}\frac{m}{r} \quad . \quad . \quad . \quad . \quad (73)$$

The Bessel functions occurring in eqns. (72) and (73) are conveniently expanded in Taylor's series about the point $h_0 s$, and to the first order of quantities we get

$$\left.\begin{aligned} J'_m(hs) &\simeq J_m(h_0 s)\left[\left(\frac{m}{h_0 s}\right)^2 - 1\right]s\delta(h) \\ J_m(hs) &\simeq J_m(h_0 s) \end{aligned}\right\} \quad . \quad . \quad (74)$$

With these approximations eqns. (72) and (73) furnish

$$\tau = -\frac{\beta_0}{h_0^2}\frac{m}{s}(jZ_z) \quad . \quad . \quad . \quad . \quad . \quad (75)$$

and
$$\delta(h) = \frac{1}{s}\frac{h_0}{k_0}\frac{\left[jZ_\phi + jZ_z\left(\frac{\beta_0}{h_0}\right)^2\left(\frac{m}{h_0 s}\right)^2\right]}{1 - \left(\frac{m}{h_0 s}\right)^2} \quad . \quad . \quad (76)$$

Thus
$$\delta(\gamma) = \frac{1}{s}\frac{h_0^2}{k_0\beta_0}\left[\frac{Z_\phi + Z_z\left(\frac{\beta_0}{h_0}\right)^2\left(\frac{m}{h_0 s}\right)^2}{1 - \left(\frac{m}{h_0 s}\right)^2}\right] \quad . \quad . \quad (77)$$

For a homogeneous and isotropic guide surface $Z_\phi = Z_z = Z_s$ and eqn. (77) leads to

$$\delta(\gamma) = \frac{1}{s}\frac{\lambda_g}{\lambda_0}\left[\frac{m^2}{(h_0 s)^2 - m^2} + \left(\frac{\lambda_0}{\lambda_c}\right)^2\right]Z_s \quad . \quad . \quad (78)$$

Thus, in this case

$$\alpha = \frac{1}{s}\frac{\lambda_g}{\lambda_0}\left[\frac{m^2}{(h_0 s)^2 - m^2} + \left(\frac{\lambda_0}{\lambda_c}\right)^2\right]R_s \quad . \quad . \quad . \quad (79)$$

with a similar expression for $\delta(\beta)$ but with X_s in place of R_s.

We observe that, although the propagation coefficients of an $H_{m,n}$-wave depend on both the circumferential and the axial components of surface impedance, the coupling coefficient τ is proportional to Z_z alone. Further, as can be seen from the above formulae, the propagation of an H_{0n}-mode is unaffected by the value of Z_ϕ; in fact the purity of H_{0n}-modes is unimpaired by the introduction of imperfect walls. Although the propagation of higher-order H-modes is influenced by Z_ϕ and Z_z, these modes are stable.

(6) APPLICATION TO RESONATORS

(6.1) General

Hitherto our attention has been focused on travelling waves; in particular we have been considering axial guided waves. By this we mean that the power as transmitted by the wave is carried principally in the axial direction, heretofore referred to as the z-direction and the assumed functional dependence was $\varepsilon^{-\gamma z}$. These waves can be conveniently denoted by $E_{m,n,\bullet}$ and $H_{m,n,\bullet}$, where $\bullet$ indicates that the wave is an axial wave.

We now proceed to investigate standing waves of the form $E_{m,n,l}$ and $H_{m,n,l}$ as met, for example, with cavity resonators. Here, as in the case of travelling waves, our field coefficients for a perfect cavity are related by

$$k_0^2 = h_0^2 + \beta_0^2 = h_0^2 + k_{z0}^2 \quad . \quad . \quad . \quad . \quad (5)$$

When the cavity walls are imperfect, however, the relevant equation is

$$k^2 = h^2 + k^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (80)$$

where, owing to imperfect side walls (walls parallel to the axis of the cavity), the coefficient h_0 becomes, as in the case of travelling waves,

$$h = h_0 + \delta h \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (81)$$

Similarly, owing to imperfect cavity end-plates (walls perpendicular to the axis of the cavity) the coefficient β_0 becomes

$$k_z = \beta_0 + \delta k_z$$
$$\beta_0 = \frac{l\pi}{L} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (82)$$

Consequently k is given by

$$k = k_0 + \delta(k) \quad . \quad . \quad . \quad . \quad . \quad . \quad (83)$$

where obviously

$$\delta(k) = \frac{h_0}{k_0}\delta(h) + \frac{\beta_0}{k_0}\delta(k_z) \quad . \quad . \quad . \quad . \quad (84)$$

Now, since

$$\left.\begin{aligned} k_0 &= \omega_0\sqrt{(\mu_0\varepsilon_0)} \\ \text{and} \qquad k &= \omega\sqrt{(\mu_0\varepsilon_0)} = (\omega_0 + \delta\omega)\sqrt{(\mu_0\varepsilon_0)} \end{aligned}\right\} \quad . \quad . \quad (85)$$

We have that

$$\frac{\delta(\omega)}{\omega_0} = \frac{\beta_0}{k_0}\frac{\delta(k_z)}{k_0} + \frac{h_0}{k_0}\frac{\delta(h)}{k_0} = \Delta_1 + j\Delta_2 \quad . \quad . \quad (86)$$

Thus, from the knowledge of $\delta(k_z)$ and $\delta(h)$ for any given cavity we can calculate from eqn. (86) the bandwidth or the Q-factor, as well as the real shift of resonant frequency from that of a perfect cavity (ω_0).

It is evident from eqn. (86) that, if we let ω_0 be the resonant frequency of a perfect cavity and ω_1 that of an imperfect cavity, then the quantity $\Delta_1 = \dfrac{\omega_1 - \omega_0}{\omega_0}$ is a function of the real parts of $\delta(k_z)$ and $\delta(h)$. The imaginary parts of $\delta(k_z)$ and $\delta(h)$ give rise to the quantity $j\Delta_2$, which has the following significance:

The resonant frequencies of any system are given, as is well known, in terms of the Q-factor of the system by*

$$\omega = \omega_0\sqrt{\left[1 - \left(\frac{1}{2Q}\right)^2\right]} + j\frac{\omega_0}{2Q} \quad . \quad . \quad . \quad (87)$$

where ω_0 is the resonant frequency of the same system in the absence of losses. If we assume that Q is a large number, the resonant frequencies will be adequately represented by

$$\omega = \omega_0 + j\frac{\omega_0}{2Q} \quad . \quad . \quad . \quad . \quad . \quad . \quad (88)$$

and, as is well known, the half-power bandwidth of the resonator is equal to the reciprocal of Q.

The comparison of eqns. (86) and (88) shows that

$$\Delta_2 = \frac{1}{2Q} = \frac{1}{2}\,\text{(bandwidth)}$$

(6.2) The Effect of Resonator End-Plate Impedance

The quantity $\delta(h)$ as deduced in Section 5 for circular guides is applicable in this case. With rectangular guides $\delta(h)$ is the sum of $\delta(k_x)$ and $\delta(k_y)$ and is applicable in this case. The quantity $\delta(k_z)$ is yet, however, to be calculated as the perturbation effect of the resonator terminating plates.

Imagine a cylindrical cavity of arbitrary shape and length L as shown in Fig. 8. Suppose now that the wall at $z = 0$ is perfect

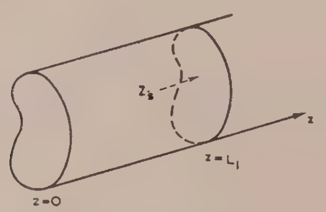


Fig. 8.—An arbitrary cylindrical cavity.

and that the wall at $z = L$ is characterized by a small surface impedance. The field components inside such a cavity (u and v are co-ordinates in the plane transverse to the axis of the guide) are:

* E.g. see PIPES, L. A.: "Applied Mathematics for Engineers and Physicists" (McGraw-Hill Book Co., Inc., New York and London, 1946), p. 164.

(*a*) *E-waves.*

$$\left.\begin{aligned} E_z &= f(u, v)\cos\beta z \\ E_u &= -\frac{k_z}{h_0^2 e_1}\frac{\partial f}{\partial u}\sin\beta z \\ E_v &= -\frac{k_z}{h_0^2 e_2}\frac{\partial f}{\partial v}\sin\beta z \\ H_u &= \frac{jk_0}{h_0^2 e_2}\frac{\partial f}{\partial v}\cos\beta z \\ H_v &= -\frac{jk_0}{h_0^2 e_1}\frac{\partial f}{\partial u}\cos\beta z \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (90)$$

(*b*) *H-waves.*

$$\left.\begin{aligned} H_z &= jf(u, v)\sin\beta z \\ E_u &= \frac{k_z}{h_0^2 e_2}\frac{\partial f}{\partial v}\sin\beta z \\ E_v &= -\frac{k_z}{h_0^2 e_1}\frac{\partial f}{\partial u}\sin\beta z \\ H_u &= \frac{j\beta}{h_0^2 e_1}\frac{\partial f}{\partial u}\cos\beta z \\ H_v &= \frac{j\beta}{h_0^2 e_2}\frac{\partial f}{\partial v}\cos\beta z \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (91)$$

Consider the effect of a homogeneous and isotropic impedance Z_s. In this case the boundary condition to be satisfied at $z = L$ is

$$\left.\frac{E_u}{H_v}\right|_{z=L} = Z_s = -\frac{E_v}{H_u} \quad . \quad . \quad . \quad . \quad . \quad (92)$$

In the case of $E_{m,n,1}$-waves $(1 \neq 0)$, if we let τ be the coupling coefficient, from eqns. (90), (91) and (92) we get

$$\frac{-\dfrac{\beta_0}{h_0^2 e_1}\dfrac{\partial f}{\partial u} + \tau\dfrac{k_0}{h_0^2 e_2}\dfrac{\partial f}{\partial v}\;\cos k_z}{-j\dfrac{k_0}{h_0^2 e_1}\dfrac{\partial f}{\partial u} + \tau\dfrac{\beta_0}{h_0^2 e_2}\dfrac{\partial f}{\partial v}\;\sin k_z} = Z_s \quad . \quad . \quad . \quad (93)$$

If the wave is reasonably pure τ will be small, and after approximating to trigonometric functions we get

$$\delta(k_z) = \frac{1}{L}\frac{k_0}{\beta_0}(jZ_s) \quad . \quad . \quad . \quad . \quad . \quad (94)$$

Similarly, for a resonator excited in H-mode

$$\delta(k_z) = \frac{1}{L}\frac{\beta_0}{k_0}(jZ_s) \quad . \quad . \quad . \quad . \quad . \quad (95)$$

The case of $E_{m,n,0}$-modes requires a little more attention. The propagation coefficients of this mode are related by

$$\begin{aligned} k_0^2 &= h_0^2 \\ \text{and} \qquad k^2 &= h^2 + \delta^2(k_z) \quad . \quad . \quad . \quad . \quad . \quad (96) \end{aligned}$$

Therefore, if the surface is homogeneous and isotropic, we obtain from eqn. (93)

$$\delta^2(k_z) = \frac{k_0}{L}(jZ_s) \quad . \quad . \quad . \quad . \quad . \quad . \quad (97)$$

(6.3) The Rectangular-Box Resonator

Consider a cavity resonator as shown in Fig. 9. Let us denote the surface impedance at $x = 0$ and $x = a$ by $Z_{x,0}$ and $Z_{x,1}$

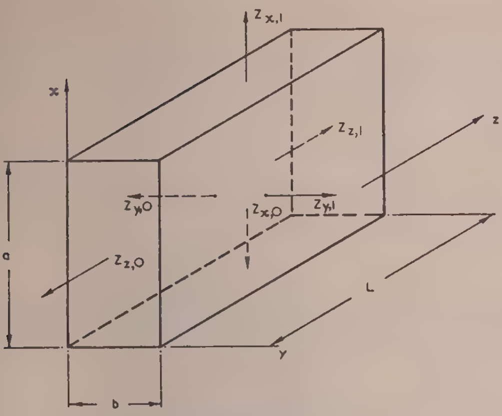


Fig. 9.—Rectangular-box resonator.

respectively, and let the surface impedance of other walls carry similar notation. If we assume that all surface impedances are homogeneous and isotropic, we get from eqn. (86) (for H_0-modes only)

$$\frac{\delta(\omega)}{\omega_0} = \left(\frac{h_0}{k_0}\right)\frac{\delta k_x}{k_0} + \frac{\delta^2(k_y)}{2k_0} + \frac{\beta_0}{k_0}\frac{\delta k_z}{k_0} \quad . \quad . \quad (98)$$

Substituting from eqns. (24), (28) and (96),

$$(\Delta_1 + j\Delta_2) = \frac{j}{2}\left[\frac{2}{ak_0}\left(\frac{h_0^2}{k_0}\right)(Z_{x,0} + Z_{x,1}) + \frac{1}{k_0 b}(Z_{y,0} + Z_{y,1}) + \frac{2}{Lk_0}\left(\frac{\beta_0}{k_0}\right)^2(Z_{z,0} + Z_{z,1})\right] \quad . \quad (99)$$

Thus the real part of eqn. (99) gives the shift of resonant frequency, while the bandwidth of the cavity is given by twice the imaginary part of eqn. (99). It is thus evident that the reactive part of surface impedance is responsible for the change in resonant frequency, while the bandwidth of the cavity depends solely on the surface resistance of the cavity walls. More precisely, inductive walls (positive X_s) bring about a decrease in resonant frequency, while capacitive walls (negative X_s) increase the resonant frequency of the cavity. Further, we note that, if $X_s = R_s$ (a metal surface), the amount by which the resonant frequency of the cavity is decreased equals one-half the bandwidth.

If all the cavity walls exhibit the same impedance, $Z_s = R_s + jX_s$,

$$\left.\begin{aligned} \frac{1}{2}\text{ bandwidth} &= \left[\frac{2}{ak_0}\left(\frac{h_0}{k_0}\right)^2 + \frac{1}{k_0 b} + \frac{2}{Lk_0}\left(\frac{\beta}{k_0}\right)^2\right]R_s \\ \text{and}\quad \Delta_1 &= \frac{1}{2}\text{ bandwidth} \times \frac{X_s}{R_s} \end{aligned}\right\} \quad (100)$$

In this particular case eqn. (100) yields the familiar expression for the Q-factor of a rectangular cavity, excited in the $H_{0,n,1}$-mode, namely

$$\frac{1}{2Q} = \pi R_s\left[\left(\frac{n}{2}\right)^2\left(\frac{\lambda_0}{a}\right)^3 + \frac{\lambda_0}{2b} + \left(\frac{l}{2}\right)^2\left(\frac{\lambda_0}{L}\right)^2\right] \quad . \quad (101)$$

where R_s is given by eqn. (8).

(6.4) The Circular Cylindrical Resonant Cavity

Consider a cylindrical cavity, as shown in Fig. 8, whose periphery is the circular radius s. Let us denote the surface impedance of the curved surface by Z_1 and that of the surfaces at $z = 0$ and $z = l$ by $Z_{z,0}$ and $Z_{z,1}$ respectively. Let us further assume these impedances to be homogeneous and isotropic.

Then, using eqn. (86) and eqns. (65) or (76) as the case may be, we get for $E_{m,n,1}$-modes

$$\left.\begin{aligned} \frac{1}{Q} &= \text{bandwidth} = \frac{2}{k_0}\left(\frac{R_1}{s} + \frac{R_{z,0} + R_{z,1}}{L}\right) \\ \Delta_1 &= -\frac{1}{k_0}\left(\frac{X_1}{s} + \frac{X_{z,0} + X_{z,1}}{L}\right) \end{aligned}\right\} \quad . \quad (102)$$

for $l \neq 0$, and

$$\left.\begin{aligned} \frac{1}{Q} &= \text{bandwidth} = \frac{2}{k_0}\left(\frac{R_1}{s} + \frac{R_{z,0} + R_{z,1}}{2L}\right) \\ \Delta_1 &= -\frac{1}{k_0}\left(\frac{X_1}{s} + \frac{X_{z,0} + X_{z,1}}{2L}\right) \end{aligned}\right\} \quad . \quad (103)$$

for $l = 0$.

For $H_{m,n,1}$-modes we get

$$\frac{1}{2Q} = \frac{1}{2}(\text{bandwidth}) = \frac{1}{k_0} \times \left\{\frac{1}{s}\left(\frac{h_0}{k_0}\right)^2 \frac{\left[1 + \left(\frac{\beta_0}{h_0}\right)^2\left(\frac{m}{h_0 s}\right)^2\right]}{\left[1 - \left(\frac{m}{h_0 s}\right)^2\right]} R_1 + \left(\frac{\beta_0}{k_0}\right)^2 \frac{R_{z,0} + R_{z,1}}{L}\right\} \quad . \quad . \quad . \quad . \quad (104)$$

while Δ_1 is given by eqn. (104) with the corresponding reactive components of surface impedances in place of the resistive components. Here, the same observations apply as have been made in connection with eqn. (99).

(7) SUMMARY OF MAIN CONCLUSIONS

A complete analysis of imperfect guides in terms of the surface impedance of the guide wall has been carried out. Particular attention was given to the rectangular and circular guides, and it was shown that all E- and H-modes are stable in circular guides, but of all the modes that could exist in a perfect rectangular guide only H_0-modes are stable.

General formulae were derived from which the propagation coefficients of imperfect guides for any small value of Z_s can be calculated. In a similar manner the bandwidth, the Q-factor and the resonant frequency of imperfect cavities can be calculated using the formulae derived.

The concept of surface impedance is useful not only where theoretical computations are involved; a design engineer equipped with a knowledge of a variety of impedance surfaces is greatly assisted by the method in the design of—for example—microwave components. Moreover, one can form a clear physical picture of many phenomena when using the impedance concept for a similar reason that one enjoys an understanding of circuit problems, provided, of course, that one becomes accustomed to thinking in terms of impedances.

(8) ACKNOWLEDGMENTS

The author wishes to express his thanks to Mr. L. Lewin of Standard Telecommunication Laboratories, Ltd., for a number of interesting and stimulating discussions.

Acknowledgment is also made to Standard Telecommunication Laboratories, Ltd., for facilities granted in the preparation of the manuscript and for permission to publish the paper.

(9) REFERENCES

(1) SCHELKUNOFF, S. A.: "The Impedance Concept and its Application to Problems of Reflection, Refraction, Shielding and Power Absorption," *Bell System Technical Journal*, 1938, **17**, p. 17.

(2) SCHELKUNOFF, S. A.: "Electromagnetic Waves" (D. Van Nostrand Co., Inc., New York).

(3) BOOKER, H. G.: "The Elements of Wave Propagation using the Impedance Concept," *Journal I.E.E.*, 1947, **94**, Part III, p. 171.

(4) KARBOWIAK, A. E.: "Theory of Composite Guides: Stratified Guides for Surface Waves," *Proceedings I.E.E.*, Paper No. 1659 R, July, 1954 (**101**, Part III, p. 238).

(5) RICE, S. O.: "Reflection of Electromagnetic Waves from Slightly Rough Surfaces," *Communications on Pure and Applied Mathematics*, 1951, **4**, p. 351.

(6) AMENT, W. S.: "Towards a Theory of Reflection by a Rough Surface," *Proceedings of the Institute of Radio Engineers*, 1953, **41**, p. 142.

(7) BUCHHOLZ, H.: "Der Einfluss der Krummung von Rechteckigen Hohlleitern auf des Phasemass ultrekurzer Wellen," *Elektrische Nechrichten-Technik*, Berlin, 1939, **16**, p. 73.

(8) BARLOW, H. M., and KARBOWIAK, A. E.: "An Experimental Investigation of the Properties of Corrugated Cylindrical Surface Waveguides," *Proceedings I.E.E.*, Paper No. 1625 R, May, 1954 (**101**, Part III, p. 182).

(10) APPENDIX

It has been shown[4] that in the case of vertical polarization, the normalized surface impedance of a perfectly conducting surface when coated with a thin layer (thickness t) of a dielectric (of dielectric constant $= \epsilon_1$) is given by

$$Z_s = (k_0 t)\left(\frac{h_1}{k_1}\right)^2 \quad . \quad . \quad . \quad . \quad . \quad (105)$$

when referred to the surface of the dielectric. Consequently, when referred to the surface of the metal, this impedance is

$$Z_s = (k_0 t)\left[\left(\frac{h_1}{k_1}\right)^2 - \left(\frac{h_0}{k_0}\right)^2\right] \quad . \quad . \quad . \quad (106)$$

In the case of horizontal polarization it is easy to show by a method similar to that illustrated in the Appendix of Reference 4 that

$$Z_s = k_0 t \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (107)$$

when referred to the plane of the dielectric. Since this is the same as the impedance of a conducting surface coated with a "thin layer of free space," the impedance as given by eqn. (107), when referred to the plane of the conductor, is zero.

DISCUSSION ON

"A GENERAL EXPERIMENTAL METHOD TO DETERMINE THE PROPERTIES OF ARTIFICIAL MEDIA AT CENTIMETRE WAVELENGTHS, APPLIED TO AN ARRAY OF PARALLEL METALLIC PLATES"*

Dr. J. B. Birks (*communicated*): The general theory of the reflection and transmission coefficients of an interface, of a symmetrical slab, and of a composite slab, which the author derives, has been given previously.† The method employed was similar to that used by the author, but the impedance concept was also introduced to extend the theory to oblique incidence and to waveguide transmission. This publication was based on work I carried out from 1942 to 1945, and it precedes the paper by Lengyel (1951) to which the author refers.

Dr. R. I. Primich (*in reply*): I am indebted to Dr. Birks for pointing out the reference to his work which had been missed. A careful study of the paper will show that the source reference for the derivation of the slab formulae is Reference 20, which lists both Dr. Birks and Dr. Redheffer, but none of these individual reports was available to me at the time. Lengyel has simply applied the formulae in Reference 20 to the system of parallel plates, and these are the formulae I have used, with a very brief outline of the derivation relegated to the Appendix. I have made no attempt to claim original derivation and I feel that a careful examination of the paper will lead back to Reference 20.

For application to artificial dielectrics the impedance concept can be misleading, and it has been very carefully and intentionally avoided, as indicated in the Introduction. The concept of complex transmission and reflection coefficients, which has been used in its stead, can include by definition the case of oblique incidence (and hence, waveguide transmission).

* PRIMICH, R. I.: Paper No. 1701 R, January, 1955 (see **102** B, p. 26).
† BIRKS, J. B.: "Dielectric Housings for Centimetre-Wave Antennae," *Journal I.E.E.*, 1946, 93, Part IIIA, p. 647.

621.396.968.08 : 621.317.341

Paper No. 1898 R
Sept. 1955

MEASUREMENTS OF THE EFFECT OF RAIN, SNOW AND FOGS ON 8·6 MM RADAR ECHOES

By N. P. ROBINSON, M.A.

(*The paper was first received* 31*st March, and in revised form* 20*th May*, 1955.)

SUMMARY

A radar set has been used to investigate the propagation of 8·6 mm radiation through various types of weather. Measurements have been made of the attenuation caused by the precipitation particles and of the intensity of radiation scattered back. No attempt has been made to correlate the results obtained with drop-size distribution, but the results are plotted against precipitation rate.

The attenuation caused by rain and the intensity of radiation scattered back by rain agreed well with theoretical predictions.

The echo intensity from fogs was too low to be detected, but the attenuation of a corner-reflector echo was measured in several fogs of different visibilities. While the results were of the same order as those predicted by theory, there was some variation, especially for fogs causing low visibility.

The attenuation caused by dry snow has not been measured, but moist snow produced attenuation two and half times greater than rain of similar precipitation rate. The echo intensity returned from the "radar bright band," composed of melting snow, was 14–19 dB more than the echo intensity from the dry snow above it, and 2–8 dB more than the echo intensity from the rain below it.

LIST OF SYMBOLS

A = Equivalent echoing area of a corner reflector.
C, C_1 = Constants.
G = Gain of an aerial over an isotropic radiator.
h = Pulse length.
M = Mean water content of a fog, g/m^3.
N = Number of drops per cubic centimetre.
p = Precipitation rate, mm/h.
P_r = Peak power received.
P_{ra} = Peak power received after attenuation.
P_{rc} = Peak power received from a corner reflector.
P_{rs} = Peak power received from a scatterer.
P_{rsh} = Peak power received from a shell of thickness δr and radius $r\theta$.
P_{rv} = Peak power received from unit volume.
P_t = Peak power transmitted.
r = Range of scatterer.
R = Range of corner reflector.
S = A scattering function (equals $\sigma/4\pi$).
V = Visual range, ft.
z = Attenuation per unit length.
Z = Attenuation, dB/km.
λ = Wavelength.
σ = Equivalent echoing area.
θ = Semi-angle of transmitted beam, measured to 3 dB point.

(1) INTRODUCTION

At wavelengths below 1 cm the effect of precipitation particles on the propagation of electromagnetic radiation is serious compared with the effect at longer wavelengths. When radiation is incident upon a particle there is some absorption of energy by the particle itself, and in addition, some of the incident energy is scattered. Both these effects produce attenuation of a beam of radiation traversing a cloud of particles. Some of the radiation is scattered backwards towards the source, and this produces the type of radar echo known as "rain clutter."

These effects have been discussed theoretically by several authors, their results being based mostly upon the theoretical computations of J. W. Ryde and D. Ryde, summarized in Reference 2. The magnitude of the effects is shown to depend upon the complex permittivity of the scatterer, the ratio of drop diameter to wavelength and the concentration of drops in the atmosphere. The permittivities chosen for water and ice were those determined by Saxton[3] for water and by Dunsmuir and Lamb[4] for ice. The numbers of raindrops per unit volume for various precipitation rates were calculated from the drop-size distributions of Laws and Parsons.[5] In the present paper the theoretical curves shown in Figs. 1 and 2 are based upon the original work of J. W. Ryde and D. Ryde.

In 1951 Saxton and Lane[6,7] published revised values for the permittivity of water. The differences between the new and earlier values are small at a wavelength of 8·6 mm, and it was not thought worth while to undertake the lengthy computations necessary to correct Ryde's scattering function *NS*. It is probable that the new values of *NS* will be at most 10% different from Ryde's values. It is simple to use the new values to calculate the attenuation due to fog, and the curve shown in Fig. 3 was drawn using the later values of permittivity. Visual range is related to attenuation, following the procedure outlined by Ryde.

There have been several measurements of attenuation, and back scatter caused by rain,[8–11] and although there is some uncertainty, none has indicated that Ryde's theory is grossly in error. No measurements in the 8 mm waveband appear to have been made, nor has the attenuation through fog been measured and reported.

Some reports of measurements in snowstorms are available,[12] and there have been measurements of the echo intensity from melting snowflakes forming the radar "bright band."[1,13,14] By considering the relative echo intensities from the regions above, below and in the bright band, the relation between the echo intensities from rain and from the snow producing it can be obtained. These observations tend to confirm Ryde's theory. Present knowledge of the echoing properties of a snowflake, of diameter comparable with the wavelength, is so inadequate as to preclude an accurate comparison between the results given in the present paper and theory.

In view of the uncertainty in some of the results published it was thought worth while to investigate the propagation of radiation in the 8 mm band.

(2) EQUIPMENT USED

(2.1) Radar Equipment

The investigation was carried out with a laboratory-built radar set using a pulsed magnetron as transmitting valve. Separate

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. Robinson is at the Radar Research Establishment, Ministry of Supply.

transmitter and receiver aerials were mounted on a frame which could be rotated by hand in azimuth, and in elevation up to 36° above the horizontal. Eighteen-inch-diameter metal searchlight mirrors fed from waveguide horns were used as reflectors. The principal characteristics of the set are listed below.

Wavelength	8·6 mm.
Transmitted peak power	15 kW (nominal, not measured).
Pulse length	0·15 microsec.
Beam width	1° 30′ (measured between 3 dB power points).
Polarization	Circular.

The signals were displayed on a range tube with a time-base of 6, 12 or 24 miles.

An r.f. attenuator, calibrated against a precision piston-type attenuator, was placed in the waveguide between the receiver horn and the input to the head amplifier. The display was calibrated using this attenuator and the echo from a corner reflector. The attenuator was also used in assessing signal strength, sufficient attenuation being inserted to bring the signal to some predetermined level.

The radar set was housed in a hut 20 ft above falling ground, giving it an uninterrupted path of about 10 miles.

(2.2) Rain Gauges

Two Mark 2 Bibby recorders and gauges were used to measure the precipitation rate at the radar site and at a site 1·75 miles distant, the instruments recording the amount of rain falling in each 1 min interval. The recorders, being separate from the collecting funnels and connected to them by land lines, were placed side by side in the same room as the radar set. Since each recorder drum was driven by a synchronous motor, time synchronization of precipitation-rate reading and radar-echo intensity could be made accurately.

(2.3) Corner Reflector

A corner reflector, mounted on a tower 65 ft above ground level 8·27 miles from the radar set, was used as a standard target. It was made to such tolerances that the equivalent echoing area of the reflector was within ½ dB of the theoretical at a wavelength of 3 mm. It had sides 70·4 cm long and an equivalent echoing area of $1{\cdot}41 \times 10^4\,\mathrm{m}^2$ at the wavelength used. The signal returned from this reflector was used to evaluate the echo intensity returned from rain and the attenuation through fogs. There was an optically clear path between the reflector and the radar set. The effective height above the immediately surrounding terrain was such that contributions to the reflector echo intensity due to an indirect path were negligible.

(3) PROPAGATION THROUGH RAIN

(3.1) Theory of Scattering by Raindrops

The intensity of radiation scattered back from rain may be calculated from the theoretical equation for the power, P_{rs}, returned to the radar receiver by a scatterer of equivalent echoing area σ, as

$$P_{rs} = P_t\frac{G^2\lambda^2\sigma}{(4\pi)^3r^4} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

The power returned from one raindrop of equivalent echoing area, σ, is obtained from Ryde's scattering function S, which by definition equals $\sigma/4\pi$. If there are N drops of equal radius per unit volume, the power returned by unit volume of the beam, P_{rv}, is given by

$$P_{rv} = \frac{P_tG^2\lambda^2 4\pi NS}{(4\pi)^3r^4} \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Rain is not composed of drops of equal radius, but by considering drop-size distributions a mean value, $\overline{NS}$, for the scattering function per unit volume may be obtained.

All drops at range r in a spherical shell of thickness δr return radiation to the radar aerial if they lie within the beam of semi-angle θ. The volume of this portion of the shell is $\pi r^2\theta^2\delta r$, provided that θ is small—a condition fulfilled in the equipment used. Hence the power returned by drops within the shell P_{rsh} is given by

$$P_{rsh} = \frac{P_tG^2\lambda^2\overline{NS}\theta^2\delta r}{16\pi r^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

The echoes returned by all drops within one-half the pulse volume will be indicated as though they were returned from drops at one range. Hence, if h is the pulse length, the power returned to the radar set during one pulse length is

$$P_r = \frac{P_tG^2\lambda^2\overline{NS}\theta^2}{16\pi}\int_{r-\frac{h}{4}}^{r+\frac{h}{4}}\frac{dr}{r^2} = \frac{P_tG^2\lambda^2\overline{NS}\theta^2h}{16\pi\left(2r^2 - \frac{h^2}{8}\right)} \quad . \quad . \quad . \quad (4)$$

For one radar set, $P_tG^2\lambda^2\theta^2h/32\pi$ is a constant, C, and $h^2/8$ can usually be neglected compared with $2r^2$.

Hence
$$P_r \simeq C\frac{\overline{NS}}{r^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

If the atmospheric and rain attenuation is z per unit length, the attenuated power returned, P_{ra}, is given by

$$P_{ra} = C\frac{\overline{NS}}{r^2}\varepsilon^{-2zr} \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

(3.2) Measurements of Attenuation in Rain

The attenuation caused by rain was calculated by measuring the echo intensity returned from the rain at different ranges at quarter-mile intervals up to 2 miles, and at half-mile intervals beyond this. From eqn. (6) it can be seen that

$$10\log(P_{ra}r^2) = 10\log C\overline{NS} - 20zr\log\varepsilon \quad . \quad . \quad (7)$$

Thus if $10\log(P_{ra}r^2)$ is plotted against r, a straight-line curve should result for rain of uniform precipitation rate. The gradient of the line is the attenuation in decibels per unit radar range, or twice the attenuation per unit range for one-way transmission. The attenuation thus obtained includes the atmospheric attenuation, taken as 0·096 dB/km at 10° C, due to oxygen and water vapour. Values of 0·02 dB/km for the former[15] and of 0·008 dB/km/g/m³ for the latter[16] are assumed, the atmosphere being considered to be saturated with water vapour at 10° C. Errors due to these assumptions are estimated at less than 0·03 dB/km, and are probably 0·01 dB/km.

To measure echo intensity the aerials were elevated 3° above the horizontal, in order to ensure that all the beam was clear of the ground. Greater angles of elevation are undesirable, as the beam may enter the bright band. The rain echo was examined on the range display, and was usually visible between the ground ray and a range of 6 miles or less, depending upon the precipitation rate. If the echo intensity changed with range in a non-uniform manner, no attempt was made to take readings. Since this was so on about 90% of all occasions when rain was falling, many erroneous results were avoided.

When the results were plotted it was sometimes found that the points did not approximate to a straight line, presumably because the precipitation rate changed with range or with time. (A

time of up to 3 min was necessary to complete a set of readings.) These results were rejected.

If the precipitation rate had been changing exponentially with range, the points still would have been on a straight line, but the rain gauges at 0 and $1\frac{3}{4}$ miles range would have read differently. If there was reason to suspect that this was so, the results were again rejected. It was found that when the precipitation rate was less than 1 mm/h the attenuation was so small that errors in the readings became comparable with the change in readings due to attenuation; therefore no results were used when the precipitation rate was less than this.

Sixty-five sets of readings were plotted for precipitation rates of 1 mm/h or more. After some had been rejected for the reasons outlined above, 51 values of attenuation from 16 storms were accepted and are shown in Fig. 1. Ryde's theoretical

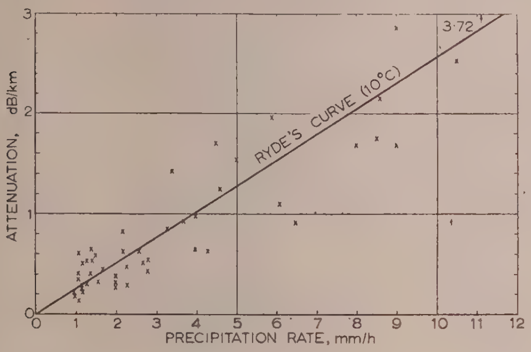


Fig. 1.—Attenuation of 8·6 mm wavelength radiation produced by rain.

curve of attenuation at 10° C plotted against precipitation rate is also shown, and it can be seen that the results are in good agreement with this. The best straight line through the experimental points was computed by the method of least squares and found to give the relationship between attenuation and precipitation rate

$$Z = 0{\cdot}26p - 0{\cdot}02$$

In Fig. 1 this would be indistinguishable from the line representing Ryde's predictions, given approximately by

$$Z = 0{\cdot}26p$$

This method of measuring the attenuation due to rain possesses some advantages over other methods, but has limitations. The chief advantage is that, by careful inspection of results, all observations made in non-uniform rain can be rejected without resorting to an excessive number of rain gauges. No careful monitoring of set performance is needed; it is essential only that

great that the rain echo would be attenuated too much to provide more than a few reliable readings. At such rates it would be better to measure the absorption over a short path length between a separate transmitter and receiver.

(3.3) Measurements of Echo Intensity from Rain

The power returned by rain at a range r is given by eqn. (4), which is

$$P_r = \frac{P_t G^2 \lambda^2 \overline{NS} \theta^2 h}{16\pi\left(2r^2 - \dfrac{h^2}{8}\right)} \simeq \frac{P_t G^2 \lambda^2 \overline{NS} \theta^2 h}{32\pi r^2}$$

when $h \ll r$. The power, P_{rc}, returned from a corner reflector of echoing area A at a range R is given by

$$P_{rc} = \frac{P_t G^2 \lambda^2 A}{(4\pi)^3 R^4}$$

Hence

$$\frac{P_r}{P_{rc}} = \frac{P_t G^2 \lambda^2 \overline{NS} \theta^2 h}{32\pi r^2} \bigg/ \frac{P_t G^2 \lambda^2 A}{(4\pi)^3 R^4}$$

$$= \frac{2\pi^2 \overline{NS} \theta^2 h R^4}{r^2 A} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Thus, with a knowledge of the appropriate values of $\overline{NS}$ and the set parameters θ and h, it is possible to compare theoretically

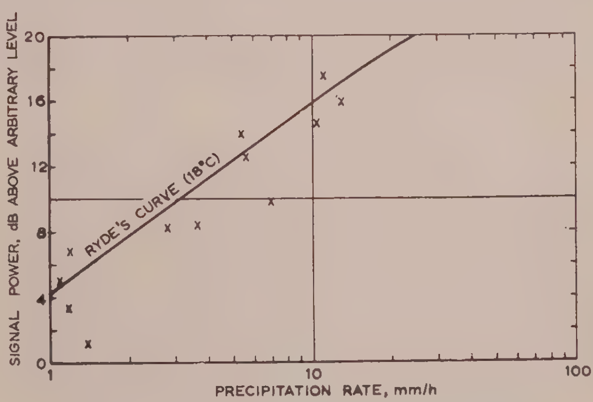


Fig. 2.—8·6 mm wavelength echo intensity from rain.

the echo intensities expected from rain at range r and a reflector at range R. The theoretical curve shown in Fig. 2 was drawn using Ryde's values for $\overline{NS}$, which are included here in Table 1.

Table 1

RYDE'S VALUES OF $\overline{NS} \times 10^8$ AT A WAVELENGTH OF 8·6 MM

Precipitation rate, mm/h	0·25	1·25	2·5	12·5	25	50
$\overline{NS} \times 10^8$	$3{\cdot}4 \times 10^{-1}$	2·2	5·0	$3{\cdot}2 \times 10$	$6{\cdot}8 \times 10$	$1{\cdot}25 \times 10^2$

the set performance should remain constant during the 3 min necessary to complete one set of readings. The method is applicable only to millimetre wavelengths, since both the attenuation and back scatter at longer wavelengths are too small. It has already been noted that values are unreliable for precipitation rates of less than 1 mm/h. It is estimated that the attenuation in rain of precipitation rate higher than 15 mm/h would be so

Fig. 2 compares the points obtained experimentally with this theoretical curve.

Readings of signal intensity received from a range of $1\frac{3}{4}$ miles were taken from the observations made during attenuation measurements. The beam of radiation had been placed so that it was above the rain gauge at a range of $1\frac{3}{4}$ miles from the radar set, and the precipitation rate recorded by this gauge was used

in the calculation of echo intensity. The radiation returned from rain was attenuated by the rain and the atmosphere, and the reading of echo intensity was corrected for 2-way attenuation, computed from the attenuation curve.

The echo intensity from the corner reflector was measured as soon as there was no rain along the path. At times several hours elapsed before this condition was fulfilled, and since the set performance may have altered somewhat, no readings of echo intensity are included for those days. A correction had to be applied to the corner-reflector reading to allow for atmospheric attenuation, and the six-hourly readings of dewpoint at Ross-on-Wye were used to calculate water-vapour absorption. These may have differed from the readings at Malvern by up to 2° F, but the possible error in the correction was small.

When the echo intensity from rain was measured, the r.f. attenuator was inserted until the signal was reduced to a predetermined level. The height of the rain echo was taken to be the maximum height reached by the echo spikes, about 5% of the taller spikes being ignored. This may be termed the "95% peak" level, and is above the r.m.s. level of echo intensity which should have been read. The corner-reflector echo, being steady, had a peak value very close to its r.m.s. value. It has been stated by Goldstein[17] that precipitation echoes have the characteristics of echoes from random scatterers, and the amplitude of the voltage vector at any instant in time follows a Rayleigh distribution. From this it can be shown that the signal level which is exceeded 5% of the time is 4·8 dB above the r.m.s. level. Thus a correction of 4·8 dB was applied to the readings of echo intensity to obtain the r.m.s. values, which were then comparable with the corner-reflector echo intensity. Since there was no accurate means of assessing whether exactly 5% of the spikes were ignored, some uncertainty exists about the value of this correction. It is felt that at least 2% and at most 10% of the spikes were ignored, and the limits of the correction are therefore 5·9 dB and 3·6 dB. Thus a maximum error of 1·2 dB may have been introduced into the readings by assuming a correction of 4·8 dB.

Examination of Fig. 2 shows that there is appreciable scatter in the results. It is to be expected that the echo intensities will display more scatter than the attenuation results, since the radiation scattered back by rain is more dependent on the drop-size distribution, differing from one rainstorm to another. The results were obtained on six different occasions between October, 1952, and February, 1953. There were many fewer readings of echo intensity than attenuation, because the corner reflector was not in position when the work began. It can be seen that the results are in good agreement with Ryde's theoretical curve, and a line drawn through the points lies about 1 dB below Ryde's curve. This discrepancy is within the limit of experimental uncertainty, which is about $2\frac{1}{2}$ dB.

(4) ATTENUATION IN FOGS

The attenuation of a beam of radiation passing through fog is of the same nature as that caused by rain, but the energy removed from the beam by scattering is proportionately much smaller, owing to the small drop size. The calculation of the attenuation through fogs is simpler than that through rain, since the drop size is such that the drop diameter can always be considered small compared with the wavelength at the frequency used.

In the calculation of the theoretical curve the treatment of Ryde[2] has been followed. The relationship between visual range and water content is taken directly from this work, which in turn is derived from a paper by Houghton and Radford,[18] who measured the drop-size distribution in fogs. Ryde concludes that for any given visual range the water content will be between $\frac{1}{2}$ and 2 times the mean water content for that visual range on 95% of all occasions. The relation between M, the mean water content, and V, the visual range, is given by

$$M = 1\,660 V^{-1\cdot 43} \quad . \quad . \quad . \quad . \quad . \quad (11)$$

Ryde shows that the attenuation is

$$\frac{4\cdot 09 M}{\lambda} C_1 \text{ decibels per kilometre} \quad . \quad . \quad . \quad (12)$$

C_1 is a constant for one wavelength and is derived from the complex permittivity of water. Using Saxton's later data,[6] a value of 0·143 has been calculated for C_1 at 20° C at the wavelength used, and the theoretical curve has been drawn using this value. C_1 changes rapidly with temperature, and at 0° C the attenuation in a fog is 1·62 times that at 20° C in a fog of the same water content. The theoretical temperature-correction factor was applied to the experimental results to make them equivalent to attenuation at 20° C. Thus any error in the theoretical variation of attenuation with temperature is included in the results.

The attenuation due to fog was measured by observing the variation of the corner-reflector response with visibility. The performance of the set was assessed by measuring the reflector response when the visibility had increased to beyond 6 000 ft, and the fog caused negligible attenuation. The atmospheric attenuation at the time of each observation was calculated assuming the atmosphere to be saturated with water vapour while fog persisted. The results are presented in Fig. 3, where attenuation is plotted against visibility.

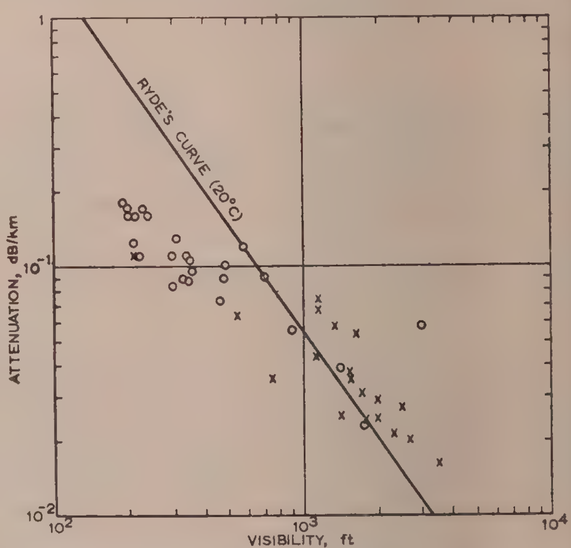


Fig. 3.—Attenuation of 8·6 mm wavelength radiation in fogs.
○○ Temperature below 0° C.
× × Temperature above 0° C.

Visibility was measured at the transmitter and a point near to the reflector. It was assessed by measuring the distance from the observer to the farthest visible large object, all measurements being made in daylight. Although this is the usual method of measuring visibility the results are not necessarily accurate. They depend upon the colour, size and background of the object, and upon the intensity and direction of the illuminating light. The two assessments of visibility often disagreed by a factor of up to 2, the value accepted being the mean of the two readings. It is certain that over the 8-mile path to the reflector there was con-

siderable variation of visibility on most occasions, producing a broad scatter of results.

At visibilities less than 500 ft the measured attenuation was less than the theoretical. The fogs causing low visibility usually occurred at temperatures below 0° C. Although supercooled water can exist well below the lowest temperature observed (−4° C), it is possible that some of the fog particles were ice crystals, which were observed visually on one occasion. The attenuation caused by an ice crystal would be less than that of a water particle of similar size. Other possible reasons for the discrepancy between practical results and theory are (*a*) an incorrect empirical relationship between visibility and the water content of fogs, (*b*) errors in the measurement of visibility, (*c*) errors in the temperature correction, and (*d*) the assumption of a saturated atmosphere may not be valid for fogs of low water content.

(5) PROPAGATION THROUGH SNOW

When the ground temperature is near or below the freezing point, precipitation usually reaches the ground in the form of snow. On these occasions propagation through snow can be studied in the same manner as propagation through rain at higher temperatures. Such occasions are rare, and during the period of observation no snow fell with the ground temperature below freezing point. When the ground temperature is well above freezing point and rain is falling, some information of the echo intensity received from snow can be obtained by elevating the beam of radiation until it passes through the snow near or above the freezing level. With this method the echo intensity from the melting band was found to be between 14 and 19 dB greater than the echo intensity from dry snow at the same range. The melting-band echo intensity was greater than that from rain by between 2 and 8 dB. As the dry snowflakes fell towards the freezing level the echo intensity often increased, presumably due to coalescence of crystals to form large flakes. The echo intensity expected from moist or dry snow at various precipitation rates can thus be estimated from Fig. 2, in which echo intensity from rain is plotted against precipitation rate.

The chief disadvantage of this type of observation is that only precipitation rate can be measured easily. Nothing is known of the type of snowflake, which is readily observed only when the snow is reaching the ground. When the ground temperature is well below freezing point the intensity of echo returned by the snow does not change with height, and so attenuation is measurable by the same method as attenuation caused by rain. When the ground temperature is near freezing point the echo intensity changes rapidly with height and the method is not applicable, since the beam must be elevated slightly above horizontal to avoid ground echoes. In four very uniform snowstorms with ground temperature between 0·4° C and 1·2° C, readings of signal intensity from the corner reflector were noted. By comparing these with the signal intensity received when no snow was present, assessments of attenuation through wet snow were made. Fig. 4 shows attenuation plotted against precipitation rate, although these readings are not as reliable as the rain-attenuation readings. The broken line is drawn through the points, ignoring the two very high readings of attenuation. These two points were obtained when the corner-reflector echo suffered some 30 dB of attenuation for about 5 min, while both rain gauges read 0·7 mm/h during and after this period. It is thought that the high attenuation was caused by an intense portion of the shower passing between the radar set and the corner reflector, but missing the gauges.

Comparison between Figs. 1 and 4 shows that moist snow produces attenuation about $2\frac{1}{2}$ times as great as that caused by rain

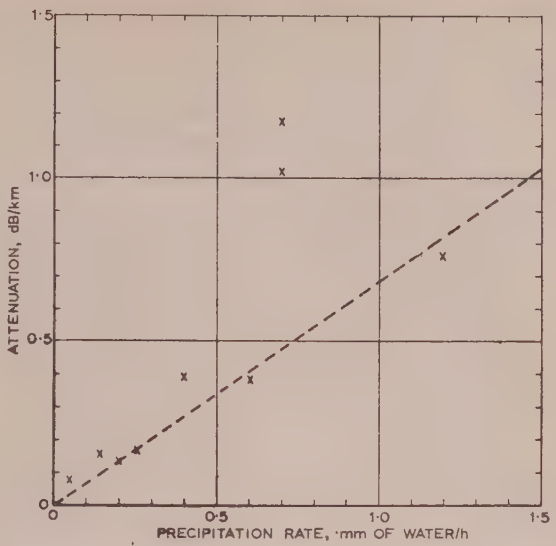


Fig. 4.—Attenuation of 8·6 mm wavelength radiation through moist snow.

of similar precipitation rate. It should not be forgotten that, owing to the change of fall velocities, the concentration of particles in the melting band is about five times as great as in the rain region below. From these few observations one would conclude that a single moist snowflake produces only about one-half the attenuation it will produce after it has turned to a raindrop.

(6) CONCLUSIONS

The experimental values of attenuation through, and intensity of radiation scattered back by, rain, snow and fog have been compared with the theoretical predictions of Ryde. The results obtained in rainstorms are in good agreement with Ryde and can be considered to be reliable. The measurements of attenuation in fogs are less satisfactory, but give values of the same order as those predicted by Ryde. A few observations of snow have been made, but owing to the many possible variations in the form of snowflakes it has not been possible to check Ryde's theory with any certainty. The echo intensity from snow is of the correct order, and the attenuation measured in most snow is about $2\frac{1}{2}$ times that produced by rain of similar precipitation rate. It should be remembered that the concentration of these particles is greater than that of the drops which results from complete melting.

(7) ACKNOWLEDGMENTS

Acknowledgment is made to the Chief Scientist, Ministry of Supply, and to the Controller of H.M. Stationery Office, for permission to publish the paper.

(8) REFERENCES

(1) Hooper, J. E. N., and Kippax, A. A.: "The Bright Band—a Phenomenon associated with Radar Echoes from Falling Rain," *Quarterly Journal of the Royal Meteorological Society*, 1950, **76**, p. 125.

(2) Ryde, J. W.: "The Attenuation and Radar Echoes produced at Centimetre Wavelengths by Various Meteorological Phenomena," "Meteorological Factors in Radio-Wave Propagation" (Physical Society, London, 1946), p. 169.

(3) SAXTON, J. A., and LANE, J. A.: "The Anomalous Dispersion of Water at Very High Radio-Frequencies—Parts I, II and III, *ibid.*, p. 278.

(4) LAMB, J.: "Measurements of the Dielectric Properties of Ice," *Transactions of the Faraday Society*, 1946, **42A**, p. 238.

(5) LAWS, J. O., and PARSONS, D. A.: "The Relation of Drop Size to Intensity," *Transactions of the American Geophysical Union*, 1943, **24**, p. 452.

(6) LANE, J. A., and SAXTON, J. A.: "Dielectric Dispersion in Pure Polar Liquids at Very High Radio-Frequencies. I. Measurements on Water, Methyl and Ethyl Alcohols," *Proceedings of the Royal Society*, A, 1952, **213**, p. 400.

(7) SAXTON, J. A.: "Dielectric Dispersion in Pure Polar Liquids at Very High Radio-Frequencies. II. Relation of Experimental Results to Theory," *ibid.*, p. 473.

(8) HOOPER, J. E. N., and KIPPAX, A. A.: "Radar Echoes from Meteorological Precipitation," *Proceedings I.E.E.*, Paper No. 846 R, January, 1950 (**97**, Part I, p. 89).

(9) ROBERTSON, S. D., and KING, A. P.: "The Effect of Rain upon the Propagation of Waves in the 1- and 3-centimetre Region," *Proceedings of the Institute of Radio Engineers*, 1946, **34**, p. 178P.

(10) MUELLER, G. E.: "Propagation of 6-millimetre Waves," *ibid.*, p. 181P.

(11) ANDERSON, L. J., DAY, J. P., FRERES, C. H., and STOKES, A. P. D.: "The Attenuation of 1·25 cm Radiation through Rain," *ibid.*, 1947, **35**, p. 351.

(12) MARSHALL, J. S., and GUNN, K. L. S.: "Measurements of Snow Parameters by Microwaves," *Journal of Meteorology*, 1952, **9**, p. 322.

(13) AUSTIN, P. M., and BEAMISH, A. C.: "A Quantitative Study of the 'Bright Band' in Radar Precipitation Echoes," *ibid.*, 1950, **7**, p. 145.

(14) BYERS, H. R., and COONS, R. D.: "The 'Bright Line' in Radar Cloud Echoes and its Probable Explanation," *ibid.*, 1947, **4**, p. 75.

(15) VAN VLECK, J. H.: "Absorption of Microwaves by Oxygen," *Physical Review*, 1947, **71**, p. 413.

(16) BECKER, G. E., and AUTLER, S. H.: "Water Vapour Absorption of Electromagnetic Radiation in the Centimetre Wavelength Range," *ibid.*, 1946, **70**, p. 300.

(17) GOLDSTEIN, H.: "Propagation of Short Waves," M.I.T. Radiation Laboratory Series, Vol. 13 (McGraw-Hill, New York, 1951), p. 551.

(18) HOUGHTON, H. G., and RADFORD, W. H.: "On the Measurement of Drop Size and Liquid Water Content in Fogs and Clouds," Papers in Physical Oceanography and Meteorology, Vol. 6, No. 4 (Cambridge, Massachusetts, 1938).

DISCUSSION ON "SUBMERGED TELEPHONE REPEATERS FOR SHALLOW WATER"* AND "THE BRITISH POST OFFICE STANDARD SUBMERGED-REPEATER SYSTEM FOR SHALLOW-WATER CABLES, WITH SPECIAL MENTION OF THE ENGLAND–NETHERLANDS SYSTEM"†

NORTH-EASTERN RADIO AND MEASUREMENTS GROUP, AT NEWCASTLE UPON TYNE, 1ST NOVEMBER, 1954

SOUTH-WESTERN SUB-CENTRE, AT TORQUAY, 17TH FEBRUARY, 1955

Mr. A. E. Twycross (*at Newcastle upon Tyne*): There are 240 components and valves per repeater, which rises to a figure of some 10 000 components and 400 valves in the transatlantic cable. This indicates the remarkable confidence now placed in these items, and reflects the detailed study given to the actual make-up of each repeater.

Valves of long life, however, are not exactly new, since some of the earlier land repeater-station valves have given 20–25 years of service with negligible degradation, but, of course, it is realized that these were of very low gain and high current consumption compared with present-day standards.

The repeater capacitance of 6 μF, and 3 μF per repeater section, would indicate a total capacitance of about 90 μF for 200 nautical miles. It seems most fortunate that a series resistor is necessary in each repeater to develop the necessary operating power, thus helping to slow down current surges when switching on or off.

Presumably the constant-current feed device tends to overcome any difference of earth potential arising from magnetic storms, etc.

The coaxial cable provides a path for the power supply and permits the use of an a.f. circuit, presumably unamplified. It would be interesting to know whether there are any peculiarities in speech reception over the a.f. path, such as noise, etc. I recently had an opportunity of observing an oscilloscope connected to one of the Scandinavian telegraph cables, and although the cable was quiescent at the time, the oscilloscope indicated much electrical activity at a very low level.

I should be interested to know the actual nature of the non-

* HALSEY, R. J., and WRIGHT, F. C.: Paper No. 1633, February, 1954 (see 101, Part I, p. 167).

† WALKER, D. C., and THOMAS, J. F. P.: Paper No. 1634, February, 1954 (see 101, Part I, p. 190). This paper was not read before the South-Western Sub-Centre.

linear device, giving a double frequency, fitted to the repeaters for testing purposes.

It is noted that the dielectric loss was five times that expected from pure polythene, owing to a tape lapping under the outer conductor. Is there any particular reason for the provision of this tape lapping?

Mr. H.M.S. Smith (*at Newcastle upon Tyne*): Mention has been made of the formation of the filiform growths on metal surfaces, and I should be glad if the authors could elaborate on this. I have had experience of such growths on silver, but have not ascertained the cause—or the cure.

The authors have stated that plastic condensers have not yet proved to be superior to paper types, and that for his purposes he prefers to use the latter. Tests which I have made of polystyrene condensers of British manufacture have shown them to be considerably more reliable than the equivalent paper types, and I should be interested to know the particular properties of plastic condensers which the authors suspect. Their reference to the failure of mica condensers with polarizing voltage was most interesting, and I should like them to elaborate on this.

The detection of a failure in one of a pair of parallel amplifiers is obviously desirable. Has any extension of this technique provided detection of failure in components in series or parallel connection?

Considerable precautions are taken to ensure that earth-return currents of appreciable magnitude do not flow in the sheath, and I note the distance of the earth electrodes from the cable. What would be the effect of submarine power transmission if the power electrodes and cable were reasonably near the telephone cable? Such conditions, presumably, apply in the Baltic Sea area.

Mr. J. E. Morley (*at Torquay*): I am familiar with the local problems arising from the maintenance of the Dartmouth–Guernsey system, which contains three repeaters in each cable, and I appreciate the problems facing the designer.

Why was it decided to use 0·62 in-diameter dielectric when, say, 0·935 in dielectric with a lower attenuation would have enabled repeaters to be dispensed with or, alternatively, would have provided more circuits—advantages which would perhaps outweigh the increased cost of the larger cable?

With reference to Section 3.2, a case can be made out for a different power-feeding arrangement, since the life of the present low-power transmitting valves CV1630 is 3–4 months only and the power supply must be interrupted before switching in the standby feeding cubicle, when the surges might be harmful to the life of the repeaters. Over one system and in one direction (Dartmouth–Guernsey) the repeaters had deteriorated by 1 dB per repeater, but not in the reverse direction. Could some form of germanium power rectifier be used? This type is now in use elsewhere, and the load required would not be more than 2 kW at 800 volts.

Could not twists get into shallow-water cables? If the repeater were allowed to "roll" these kinks would come out, but it is the intention to allow the repeater casing to remain rigid on the sea bed.

It has been stated that usually the components and not the valves fail, and yet it was recorded that the failure of more than one capacitor was due to damage in proof testing. Does this indicate too-stringent proof testing?

Messrs. R. J. Halsey, F. C. Wright, D. C. Walker, and **J. F. P. Thomas** (*in reply*): In reply to Mr. Twycross, the constant-current power supply does overcome the potential differences due to earth currents, but shallow-water systems are usually short in length and these potential differences are small compared with the power-unit e.m.f. However, on longer systems (e.g. transatlantic) they are important, and the power units must be capable of regulating over a wider range of line voltage. No trouble from magnetic storms has so far been experienced by the British Post Office on a.f. circuits, not even on the Lowestoft–Borkum route which is in a similar geographical position to the Scandinavian telegraph cables quoted. The non-linear device for frequency doubling contains germanium-crystal diodes—see Section 3.9 of the paper by Roche and Roe.* Polythene cables with extra insulating-tape lapping are now obsolete, but it was presumably originally added to reduce the risk of damage to the polythene by the outer conductor tapes.

The filamentary growths referred to on metal surfaces are of pure metal, and tin is one of the worst offenders; Mr. Smith is referred to Reference 18 of the paper by Halsey and Wright for further details. However, metallic growths on silver usually occur only in the presence of an electric field, but very often atmospheric conditions induce growth of insulating filaments of a silver salt (e.g. silver sulphide). It is realized that plastic capacitors are now extremely reliable, but being relatively new, they have not the proven integrity of paper capacitors for the long life required of submerged repeaters. The liability of mica capacitors to fail under polarization has been known for some time, but the mechanism of failure is still not yet understood—it is probably caused by foreign material in the mica stack or within the mica itself.

In answer to Mr. Morley's question on the Dartmouth–Guernsey system, the repeaters were inserted in existing 0·62 in paragutta cables; even if new cables had had to be laid, a larger cable capable of providing 60 circuits without repeaters would have proved a more expensive scheme on capital costs. With regard to his comments on the power supplies, the CV1630 valve is the constant-current control valve and not a rectifier; there would be little advantage in replacing the present metal rectifiers with a modern germanium type, since a relatively low efficiency is inherent in a stabilized power unit. Experience has shown that the life of CV1630 valves is of the order of a year, and not 3–4 months as stated by Mr. Morley, and so interruptions of the power supply are infrequent. It is appreciated that it is desirable not to subject repeaters to power switching surges, but this is necessary for testing purposes, and there is no evidence to prove that the repeater life is affected. The "deterioration" referred to by Mr. Morley has not been confirmed, and the order of the change quoted could well be accounted for by seasonal temperature variations which, of course, are more apparent in the high-frequency band.

The anti-rolling plates were put on repeaters to cover cases where they might be laid on an inclined sea bed. Omission would not prevent twists in the cable, and no such trouble has been experienced when the repeaters have been carefully laid.

The amount of testing on components is a difficult problem, since the extra handling and testing may affect their life. It is agreed that proof testing of capacitors, for example, can be dangerous, but a low-voltage test for a short period is useless in eliminating the defective capacitors. Ionization tests are not always practical, and when applicable, the results are often misleading.

* Roche, A. H., and Roe, F. O.: "The Netherlands–Denmark Submerged-Repeater System," *Proceedings I.E.E.*, Paper No. 1635, February, 1954 (**101**, Part I, p. 180).

DISCUSSION ON
"A SHORT MODERN REVIEW OF FUNDAMENTAL ELECTROMAGNETIC THEORY"*

MERSEY AND NORTH WALES CENTRE, AT LIVERPOOL, 29TH NOVEMBER, 1954
NORTH-EASTERN RADIO AND MEASUREMENTS GROUP, AT NEWCASTLE UPON TYNE, 21ST MARCH, 1955

Mr. J. E. Macfarlane (*at Liverpool*): If a simple loop is rotated at a uniform speed in a uniform magnetic field and the circuit is complete, a current will flow. This early introduction to electromagnetic theory is best considered as change of flux linkages, the loop being "filled" twice and "emptied" twice in one revolution, i.e. four changes.

Another early demonstration is to plunge a permanent magnet into a solenoid coil connected to a galvanometer. Then, by the principle of work done, the direction of the galvanometer deflection can be both deduced and demonstrated. The demonstration should then be repeated with an electromagnet replacing the permanent magnet, both without and with an iron core. The flux-cutting idea should not be introduced until after these experiments.

M.K.S. units will eventually make things easier for the students. We found last session that a class of civil engineers taking S2 Electrical Engineering Science for the first time accepted these units, and the examination results were satisfactory; but as this is the first session during which M.K.S. units will be in use throughout Liverpool technical colleges, a few years' experience will be needed to show the overall effect.

Mr. D. Chalmers (*at Liverpool*): Until the introduction of M.K.S. units my starting point when teaching electromagnetic theory was to use the expression for the force between two magnetic poles, $F = m_1m_2/\mu d^2$ dynes, in C.G.S. units. From this expression μ can be more or less determined and a whole system of magnetic and electrical units constructed.

Since the suggested use of the M.K.S. system I have modified the teaching sequence; for example, it is possible to show parallel concepts in the conduction, magnetic and electrostatic circuits as follow:

Conduction circuit.—Current density = conductivity × voltage gradient.

Magnetic circuit.—Magnetic flux density = permeability × magnetizing force.

Electric flux density.—Permittivity × electric field intensity.

The three circuits have three analogous quantities: (*a*) current density, magnetic flux density and electric flux density, which can be regarded as the effect; (*b*) conductivity, permeability and permittivity, regarded as the associated material constants; (*c*) voltage gradient, magnetizing force and electric field intensity, regarded as the cause. Admitting that there are in the magnetic and electrostatic cases two unknown quantities, permeability and permittivity, a rational system of units can then be established using the M.K.S. system.

From my teaching experience the M.K.S. system has been favourably accepted by students, although in first- and second-year students in technical colleges the appreciation of the absolute permeability of $\mu_0 = 4\pi \times 10^{-7}$ Wb/m² per AT/m is somewhat vague.

Mr. E. D. Taylor (*at Newcastle upon Tyne*): There will be little opposition to the proposal that the electrical engineering student should be given a logical presentation of the fundamentals of electromagnetism; indeed, if he is worth his salt he should demand it for himself and will choose the presentation that seems the most satisfactory. Another person to whom the presentation is important is the man in industry who has to solve individual electromagnetic problems, perhaps for design purposes or for explaining phenomena. His needs are not merely a logical presentation, but practice in applying and thinking around the basic equations, not in the sense of mathematical analysis, but in the sense of understanding what are assumptions, what are deductions, what are experimental data and what are merely convenient ways of looking at things, such as, for example, Fig. 1 and its explanation. I suggest that more could be done in setting students some of the problems that are illustrated in Figs. 4, 5, 6 and 9. Besides these, there are other simpler and more common ones in shielding, short-circuited turns and so on which require considerable clarity of thought.

It is questionable whether the paper really displays a completely logical sequence of clear thought; for example, in the ready acceptance of the vectors $\boldsymbol{B}$ and $\boldsymbol{D}$ in eqns. (9) and (10) and rejection of vector $\boldsymbol{A}$ in eqn. (15), particularly since vector potential is such a powerful tool in later work besides helping to avoid much of the confusion met with by the statement of Neumann's law. Similarly, a mind which appreciates the symmetry of equations will surely ask the question that leads to $\boldsymbol{A}$, namely: "$\boldsymbol{B}$ being a vector, of what could it be the derivative?" Again, current elements are not favoured as being unrealizable, although there is ready acceptance of Fig. 1(*b*), which is equally unrealizable.

With respect to the symmetry of the equations, mathematical symmetry is, after all, a subtle point. It certainly has to be forced on the equations with the need for magnetic conduction-current density (= 0). Asymmetry itself is valuable in drawing attention. Here it makes us pay attention to the first and simplest electromagnetic observation, namely that the magnetic field, unlike the electric field, has zero divergence, i.e. when the magnetic field is due to currents, there are no sources of magnetic flux which correspond to electric charges as sources of electric flux. In any case, the equations are really no easier to remember, and the need for displacement current, if not all the reasoning leading up to it, is clear by trying to omit it from eqn. (38); div. $\boldsymbol{i}$ is then zero, which is absurd.

Mr. G. H. Hickling (*at Newcastle upon Tyne*): The Faraday-disc experiment, and the variations on this of the rotating magnet generator and the homopolar motor, serve to illustrate circumstances in which only the flux-cutting rule for the generation of e.m.f. (and the inverse effect in the last example) is applicable. The more important deduction to be drawn from these experimental demonstrations, however, concerns the manner in which the electric circuit and the magnetic field must be defined in applying these rules. In each of these instances the geometrical form and position of the electric circuit is kept

* HAMMOND, P.: *Proceedings I.E.E.*, Paper No. 1598, December, 1953 (see 101, Part I, p. 147).

fixed, as is the path in which the current flows; but in each instance a part of the metallic conductor comprising the circuit is in motion in the field, carrying with it the charges on which the electromagnetic forces consequently act. A further familiar example of this is the eddy-current disc in the induction watt-hour meter.

Equally significant is the deduction that the magnetic field due to the bar magnet rotating about its axis remains at rest, even inside the magnet itself. The view to be taken here, evidently, is that the field is a mass effect due to many molecular magnets (or electron spins) and remains unaffected so long as the distribution of these in the region considered is not disturbed.

Concerning the general thesis of the paper, I would fully concur with the author's wish to develop the electrostatic and electromagnetic series of units in symmetrical form from the inverse-square law of forces on the unit charge and unit pole, the two systems being linked ultimately by the connecting constant of 3×10^{10} cm/sec. It is unfortunate, however, that this is linked in the paper with the application of vector notation, which, however useful, is unfamiliar to the elementary student. Furthermore, I believe that the development of the unit systems is very much clarified by defining the unit of force as 1 dyne in eqns. (1) and (2), rendering the constants α and β unnecessary.

Mr. A. O. Carter also contributed to the discussion at Newcastle upon Tyne.

Mr. P. Hammond (*in reply*): Although the paper dealt only in passing with the M.K.S. system of units, much of the discussion has centred on the issues arising from the introduction of this system into electrical engineering courses. It seems that to many speakers the alteration in the size of the basic units implies also a change in the sequence of instruction that is to be used. Because the units are relatively new, and hence progressive, we are urged to embrace also the new and progressive sequence of instruction. I still hold the opinion expressed in Part 1 of the paper. A mere shifting about of factors such as 4π or 10^8 can never contribute to a clearer understanding of electrical science. To believe with Mr. Macfarlane that students gain an understanding of electrical matters more easily in M.K.S. units is almost equivalent to a belief that we should all be richer if we had a decimal system of currency. Mr. Chalmers rightly points out what is in store for us teachers when he says, in a masterly understatement, that the appreciation of certain students of the nature of absolute permeability is somewhat vague. There is no doubt that the vagueness arises largely from a return in teaching to the view of Gilbert (*c.* A.D. 1600), who thought of electricity in terms of orbs of virtue surrounding electrified bodies.

I find myself in almost complete agreement with Mr. Taylor. The vector potential is far too useful a tool to be dismissed lightly. It might well be argued that the delayed potentials are more useful to the electrical engineer than Maxwell's equations, because the potentials already include the boundary conditions of the problem. Nevertheless, it seems to me that a clearer picture of electromagnetic radiation is conveyed by a sequence of instruction that starts with electric and magnetic charges, goes on to the forces exerted by these charges and only then introduces the idea of potential as a convenient device used in the summation of forces. I am grateful to Mr. Hickling for his clear statement of the flux-cutting effect and of the action of a rotating magnet. Vector algebra is certainly not necessary for the presentation of the matter contained in the paper. We use it only with students reading for Part II of the Mechanical Sciences Tripos.

DISCUSSION ON "A TRANSATLANTIC TELEPHONE CABLE"*

NORTH-WESTERN CENTRE, AT MANCHESTER, 1ST FEBRUARY, 1955

Mr. G. R. Polgreen: Before the submerged repeater had been developed much work was done in the continuous loading of submarine cables, and the paper describes the project of 1928 for a single telephone channel loaded with Perminvar. My limited experience with this material was that it had very low hysteresis loss—essential for reducing cross-modulation to a minimum—but it was somewhat unstable. It would be useful to know whether this instability was in any way a limiting factor. Just before the war I was engaged on experimental work of loading a submarine cable with compressed rings of insulated nickel–iron alloy powder which were either pre-moulded on to the centre conductor of a coaxial cable at intervals or slipped on continuously as sleeves. The leakage reactance appeared to be adequate but the hysteresis loss was somewhat high. This subject was also briefly mentioned in a paper† on the possible applications of ferrites. Do the authors think that this system has advantages for submarine telephone cables?

Mr. H. G. Davis: If the venture is successful it will probably be found that the capacity of the new cables will soon be strained to the limit. In this connection, Dr. Buckley in the 1942 Kelvin Lecture indicated that 40 circuits was at that time a conservative estimate of requirements, and even 100 circuits might not be found excessive. Against a demand of this order, the 36 circuits and 100 kc/s frequency band of the American designed system may soon prove inadequate. By contrast, the British design, using 2-way amplifiers at 20 nauts spacing and transmitting 240 kc/s in each direction, makes the greater appeal; but the British rigid repeaters have not yet been proved at depths beyond 1 700 fathoms.

The success of the scheme depends equally on the cable and the submerged repeaters. Much attention has been focused on the performance and behaviour of the repeaters, but it may well be that cable performance and life are the decisive factors. The behaviour of the cable itself is vitally important for two reasons. First, as indicated in Section 8.4, repair repeaters will probably have to be inserted after each deep sea repair, to compensate for the additional length of cable that must necessarily be inserted. Secondly, in Section 10 it is stated that a margin of only 250 volts is allowed at the cable terminals for supplying power to any additional repeaters inserted. This corresponds to four or five additional repeaters, and, based on an economic life of 20 years, this only allows for one deep-sea repair every four or five years. Have the authors evidence to confirm that this fault rate is unlikely to be exceeded under transatlantic conditions?

* Kelly, Mervin J., Radley, Sir Gordon, Gilman, G. W., and Halsey, R. J.: *Proceedings I.E.E.*, Paper No. 1741, September, 1954 (see **102** B, p. 117).

† Latimer, K. E., and Macdonald, H. B.: "A Survey of the Possible Applications of Ferrites," *Proceedings I.E.E.*, Paper No. 900 M, October, 1949 (**97**, Part II, p. 257).

What of future developments? Generally speaking, a major item in the cost of small submarine cables is the armouring. It is therefore essential that a given size of cable shall be made to carry the maximum number of wide-band repeaters that can be supplied with power over the cable itself. Section 13 stresses the possibilities of transistors in this connection. If the ruggedness and long life claimed for these devices can be established under deep-sea conditions, this may very well be the direction in which future development will take place.

Mr. C. F. Campbell: Is it proposed to control the performance of the circuit by means of pilot signals? I assume that the basic equalization is included in the feedback circuit of the submarine repeaters, but are deviation or temperature equalizers envisaged? Will any attempt be made to control the gain of repeaters, or will all control be exercised at the terminals?

Failure of a valve heater will result in a break in the power-feeding circuit. This break will be closed by the gas diodes fitted in all the submarine repeaters; but since these diodes have a finite ionization time, is it not possible that a voltage surge will travel in both directions along the cable from the faulty repeater, and so cause breakdown of neighbouring repeaters?

It is intended to pass about 0·25 amp d.c. down the centre conductor of the cable, and to use the sea itself as a return path. Is there no danger of the return currents flowing into and out of the armouring and outer conductor of the cable, and so causing electrolytic corrosion?

Finally, it seems probable that traffic will be at a minimum on the cable between about 0400 and 0700 hours G.M.T. Have any schemes been considered for earning revenue during this time? The possibilities which come most easily to mind are high-speed telegraphy, facsimile transmission, recorded radio programmes and recorded greetings.

Mr. H. Wood: No mention is made of any provision for broadcast circuits, although the system will presumably handle the standard carrier broadcast channel recommended by the C.C.I.F. When one considers the quality of broadcast transmission which can be achieved over the proposed system and combines that with the new frequency-modulated radio transmission of the B.B.C., the absence of noise in future broadcasts from the western hemisphere should be quite striking when we compare it with what we have come to regard as the normal standard.

The simplification of the terminal equipment arising from the decision to make the lowest frequency 20 instead of 24 kc/s over the British Post Office section is not at all obvious. It appears that the low-frequency group is produced by a single stage of modulation from a partially equipped supergroup, and thus to change the bottom frequency by 4 kc/s means only a change in the carrier frequency by this amount.

It would appear that the arrangements for feeding power to the repeaters in the Atlantic section are the same for both cables, i.e. the direction of the flow of current is the same in both cases. What considerations led to this arrangement rather than one in which the direction of the flow of current is different in the two cables? This latter arrangement would appear to offer some advantage in reducing the corrosion problem, in that if the system were completely balanced, there would be no current flowing through the sea at all.

Dr. W. J. J. Curry: Section 8.1 states that "impedance irregularities must also be avoided, since they affect the gain/frequency characteristics of the repeater." What magnitude of impedance irregularity do the authors consider practically significant in relation to the repeater requirements?

On other non-submarine types of coaxial cable the pulse-echo method of measuring and locating impedance irregularities has found favour, not only as a means of test, but also as a method of examining the cable in its various stages of manufacture and during installation. This has facilitated improved designs of cable which are less susceptible to damage under normal conditions of installation. Do the authors consider the method a useful one in relation to submarine-cable development, and, in particular, may it contribute to our knowledge of cable characteristics during laying operations?

Mr. J. F. Shipley: I was interested to learn that, to prevent kinking while laying or when standing-to, the submarine cable must be kept in tension. This coincides with my knowledge of towing floating docks, with which vessels I am often concerned. These unwieldy ships, when towed across the Indian Ocean, for example, must avoid the monsoon. If the towing time is too early the dock has to be towed round in circles of several miles diameter, depending upon the time to be wasted before continuing the voyage. During the whole of this time the tow-ropes must be kept taut or the dock may get out of control, which might lead to its loss at the outset of its possible life.

Mr. J. E. L. Robinson: The comparison between the American more mechanically advanced, although electrically more conservative, deep-sea section of the equipment and the British shallow-water gear is most interesting. It is suggested that the improved communication provided by this cable is likely to lead to a rapidly rising demand for more telephone channels than the cable can provide. With this possibility in mind, what design will be adopted for a deep-sea section to be installed in, say, 5 years' time? It would seem that germanium amplifiers will not be ready. Will the authors then use throughout the more modern higher-mutual-conductance valves and a 2-way cable, as in the present British equipment?

Since each terminal generator is stated to be capable of supplying the load alone and also has 20% overvoltage capacity, it seems to follow that, with one generator operating normally and providing half the load, maximum voltage accidentally generated on the other machine would impose 1·7 times the normal voltage on the filaments. And if there were collusion by evil persons at both ends of the cable, the overvoltage could be 2·4 times normal. Under these conditions presumably only one filament is likely to fail, but the others might well be so damaged that the replacement of every repeater would be necessary. Is this costly contingency well guarded against?

Mr. R. W. Presswell: If the gland G in Fig. 9 were moulded in a harder grade of polythene, would it be used with confidence at ocean depths?

Is the adhesion between the polythene moulding and metal castellated boss of the gland adequately strong and reproducible?

Since a quantity of sea water may leak into the annular space between the pressure resisting housing and the inner sealed cylinder A without affecting the performance of the amplifier, would there be any justification in replacing the silver-solder seat between the bulkheads and casing by a mechanical seal based on the O-ring principle?

If a second transatlantic telephone cable is to be laid, would it be operated on the one-way or two-way system?

Dr. M. J. Kelly, Sir Gordon Radley, Mr. G. W. Gilman and **Mr. R. J. Halsey** (*in reply*): It will help to clear up some misunderstandings as to the relative merits of the American and British systems to emphasize that on the main crossing the capacity of the system is limited by the power circuit. The safe voltage for the system, divided by the voltage required by each repeater, determines the number of repeaters which can be operated and hence the maximum gain which can be used and the maximum frequency which can be transmitted. Neither 2-way repeaters nor higher-performance valves could increase the circuit provision over this section; only reduced voltage require-

ıents for each repeater could do this—hence one attraction of ·ansistors for the future.

On this main section power is fed in opposite directions over ıe two cables, and so there is a negligible earth current. On the larenville–Sydney Mines section, where there is only one cable, ıere are long earth cables designed to keep low the current eturning via the cable sheath. Mr. Robinson's calculations of xcess voltage on valve heaters are not valid, since constant-urrent generators are used. Experience with telegraph cables hows that deep-water-cable faults are very rare, and we can easonably expect that not more than one such fault will occur ı the first ten years.

There is no variable gain or equalization at any submerged epeater, adjustments being made manually at the shore stations ı accordance with the indications of pilot signals. The gain tability of the system with variations of ocean temperature will e adequate to allow this. There are a number of pilots for ection and overall control.

The cable specification for the main crossing requires that npedance irregularities shall be such that the echo attenuation t any point, measured with a 1·5 microsec sine-squared pulse, hall be at least 45 db. Since the repeaters have high input and utput impedances they are very sensitive to impedance irregu-ırities. Pulse-echo tests are carried out at various stages in the ıanufacture of the cable, but they will not be used during the ıying operation as the (one-way) repeaters will be inserted in dvance and therefore cannot transmit the reflected signals.

The use of a 20 kc/s lowest frequency on both systems allows ıe same supergroup translating frequency to be used on both ystems at Clarenville, minimizes crosstalk problems at that point and allows space for supervisory signals immediately above the low-frequency transmission band of the British system. Music channels will be provided.

The grade of polythene used for the cables is 0·3, and the same grade is used for the glands on the British repeaters. These glands are regularly tested at 5 tons/in^2 and will withstand this pressure for very long periods without failure or change. The bond is certainly effective, and its efficacy is reflected in a greatly increased resistance to a torque tending to turn the gland on the castellations. There appears to be no reason why O-ring seals should not be effective, although experience with them in this connection is comparatively short.

Proposals to apply continuous loading for high-frequency cables look attractive, but really satisfactory designs never seem to emerge; loading by means of ferrites would probably have theoretical advantages. For ocean telephony the simplicity and uniformity of unloaded cable have great merit, but it would be rash to predict that there is no future for some form of continuous loading; this would have to be applied in such a way that no voids are left around the inner conductor. Regarding the use of Perminvar in the 1928 project; the instability referred to by Mr. Polgreen was definitely not a limiting factor; the Perminvar was somewhat unstable during the laying operation, but remained quite stable after laying.

The possibility, referred to by Mr. Campbell, that the operation of a gas diode in one repeater will cause a voltage surge which will, in turn, cause a breakdown in neighbouring repeaters has been investigated. The results show that the attenuation between repeaters reduces the surge below the value required to operate the gas diodes in the adjacent repeaters.

DIGEST OF A PAPER

CURRENT SUMMATIONS WITH CURRENT TRANSFORMERS

621.314.224 : 621.317.31 Paper No. 1678 M

A. HOBSON, M.Sc.Tech., Associate Member

(DIGEST *of a paper published in June, 1954, and to be republished in 1955, in Part A of the* PROCEEDINGS.)

A prejudice exists against using current transformers for otalizing the currents in a number of supply lines because it is hought that the accuracy is impaired if one or more lines are idle. The present paper was begun with the idea of showing that the rejudice is based on a fallacy, and that the errors are stable vithin close limits, but the consequent simplification of the circuit arameters has enabled formulae to be derived for estimating the ffective burden on each transformer and also for computing the verall error from tests on the individual units.

The object of summation is to enable a number of currents in lifferent feeders to be regarded as a single current. It is therefore s a single current transformer with a definite overall ratio of its wn that a complete summation transformer equipment should e considered to act, its performance being judged on that asis. For this to be so the overall error should depend only n the total load and not on the way it is shared among the eeders.

Two methods of summation are in general use, both requiring separate current transformer in each line. The first method is o connect all the secondaries in parallel to the measuring burden, vhich carries the sum of their currents. It is essential for all he transformers to have the same ratio, and the current in the burden may be quite large if standard 5 amp secondaries are used.

It is generally preferable to use a summation current transformer to sum the secondary currents. This has as many primaries as there are feeders. Each primary is connected to an appropriate line transformer secondary, and its own secondary provides a current proportional to the total load.

The summation transformer is more flexible than direct paralleling of the secondaries, and will handle not only unequal transformer ratios but also unequal secondary currents. Moreover, the current in any part of the secondary circuits need not exceed 5 amp, enabling standard current transformers and instruments to be used.

IDLE FEEDERS

Here the behaviour of current transformers in parallel will be studied for different conditions of load sharing. Similar arguments apply to circuits employing a summation transformer.

In Fig. 1 only the left-hand current transformer is loaded, and it is clear that part of its output is diverted from the burden to provide a magnetizing current for the idle transformer. If the latter were disconnected the accuracy of measurement would clearly be improved, and it is this fact which causes the doubts as to the validity of the method.

Mr. Hobson is with Smith Hobson, Ltd.

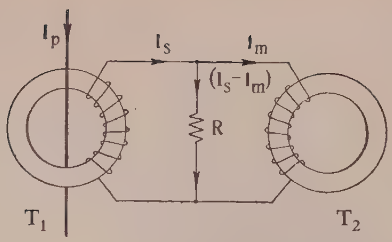


Fig. 1.—Parallel current transformers with one line idle.

That is not, however, a complete approach to the problem. Current transformer secondaries are not automatically disconnected when their primaries cease to carry load, so that their magnetizing currents are present all the time. What is really important is that the overall accuracy should be stable for a given total load, and should not vary according to the way in which the current is shared between the lines.

Now the total error is the sum of the magnetizing currents of the individual transformers, which depend on the voltages induced in their secondaries. If the impedances of the windings and leads are ignored, exactly the same voltage exists across both secondaries. This depends only on the burden current, which in turn is proportional to the total load current and is not affected by its distribution. For a given total load, therefore, the two magnetizing currents, and consequently the overall error, have fixed values, and this applies even when all the current is in one feeder.

When the impedances of the windings and leads are taken into account it can be shown that for a given total load the average e.m.f. induced in the secondaries is always the same, and if the magnetization curve of the cores were linear the overall erro would be quite stable. In practice this is not so and sma variations occur owing to non-linear permeability.

Test Results

The stability of the overall error is amply confirmed in Table which gives test figures on four similar 100/5 amp ring tran

TABLE 1

Effect of Idle Feeders on Overall Error

Burden current	Equally loaded feeders		One idle feeder		Two idle feeders		Three idle feeders	
	Ratio error	Phase angle	Ratio error	Phase angle	Ratio error	Phase angle	Ratio error	Phas angl
amp	%	min	%	min	%	min	%	min
6	−0·83	27	−0·83	27	−0·81	27	−0·78	27
5	−0·82	30	−0·83	29	−0·81	29	−0·78	29
4	−0·81	33	−0·81	33	−0·80	32	−0·76	32
3	−0·79	36	−0·79	36	−0·77	35	−0·74	35
2	−0·76	41	−0·76	40	−0·75	40	−0·72	39
1	−0·72	47	−0·72	46	−0·70	45	−0·67	45

formers connected in parallel to a burden of 0·1 ohm. Th overall ratio was thus 400/20 amp and the burden 40 watts rated current. Transformers with larger errors than usual we chosen so as to magnify the differences. The measurements we made on an ordinary current-transformer testing set, and th maximum burden that could be handled was 6 amp.

Since it was difficult to arrange four identical feeders in parall a special test procedure was adopted. The results in the fir

TABLE 2

Unbalanced Transformers in Parallel

Rated current	Individual errors				Errors when in parallel					
	No. 1		No. 2		Equal loads		No. 2 idle		No. 1 idle	
	Ratio error	Phase angle	Ratio error	Phase angle	Ratio error	Phase angle	Phase angle	Phase angle	Ratio error	Phase angle
%	%	min	%	min	%	min	%	min	%	min
60	−1·3	35	−0·56	17	−0·95	26	−1·0	26	−0·90	26
40	−1·5	50	−0·56	20	−1·05	35	−1·1	35	−1·0	34
20	−1·8	80	−0·53	25	−1·2	54	−1·3	55	−1·1	51
10	−2·15	125	−0·50	30	−1·35	80	−1·45	83	−1·2	83

TABLE 3

Effect of Different Amounts of Turns Compensation

Rated current	Individual errors				Errors when in parallel					
	No. 1		No. 2		Equal loads		No. 2 idle		No. 1 idle	
	Ratio error	Phase angle	Ratio error	Phase angle	Ratio error	Phase angle	Ratio error	Phase angle	Ratio error	Phase angle
%	%	min	%	min	%	min	%	min	%	min
60	−0·05	36	−0·56	17	−0·3	27	+0·1	28	−1·05	28
40	−0·25	50	−0·56	20	−0·45	35	0	36	−1·05	35
20	−0·65	83	−0·53	25	−0·63	55	−0·1	55	−1·15	53
10	−1·05	130	−0·50	30	−0·80	82	−0·35	86	−1·25	83

column are for equally distributed load and were obtained with the four primaries in series, giving a test ratio of 100/20 amp. For the second column the primary cable was removed from one transformer, leaving its secondary idle, and the test ratio was then 100/15 amp. The third and fourth columns were similarly obtained by having two and three idle transformers, but each time the tests were made at the same burden currents, representing the same total load.

Table 2 gives similar results for two very unequal transformers in parallel, and comparison of the results in the last three columns shows that even under these severe conditions there is little change in error between the extremes of load sharing.

More important is the effect of having different amounts of turn compensation on the various transformers. Table 3 refers to the same transformers as Table 2, with the secondary winding of the first unit reduced from 80 to 79 turns. The ratio error of this transformer, given in the first column, is seen to have changed by about 1·25% in a positive direction, as would be expected.

The ratio error for distributed load, col. 3, has changed only by about half of 1·25% as compared with Table 2, since only half the current comes from the biased transformer. When the good transformer is idle, all the current comes from the biased unit and the ratio error is made still more positive. When the poorer one is idle, its turn compensation does not come into action at all, yet its full magnetizing current has still to be supplied, and the ratio error shifts negatively by about half the bias percentage from the mean position. The conditions are then very nearly the same as for the final column of Table 2, with which the last column in Table 3 may be compared.

Position of Parallel Connection

It is commonly thought that the secondaries must be paralleled at the burden and not at the transformer terminals. The reason given is that the voltage across the idle secondaries is thereby reduced by the drop which would otherwise occur in the common leads, causing a smaller magnetizing current to be diverted from the burden. Since it has been shown that idle feeders have no appreciable effect, it follows that the actual point of paralleling is unimportant and may be chosen according to convenience.

In the remainder of the paper formulae are derived for estimating the effective burden on each transformer and also for computing the overall error from tests on the individual units.

DIGESTS OF INSTITUTION MONOGRAPHS

THE FREQUENCY-RESPONSE ANALYSIS OF NON-LINEAR SYSTEMS

621–526 : 621.373.421 Monograph No. 126 M

P. E. W. GRENSTED, M.A.

(Digest *of a paper published in April,* 1955, *as an* Institution Monograph *and republished in September,* 1955, *in Part C of the* Proceedings.

If a non-linear system is oscillating, either through external excitation or internal regenerative action, it may be possible to show that the waveform at the inputs to the non-linear elements in the system is approximately sinusoidal. The frequency-response analysis of the system as a whole may then be carried out on the assumption that all harmonic components generated by the non-linear elements can be ignored.

This principle has been applied to the analysis of "on-off" control systems by Kochenburger,[2] of non-linear control systems in general by Johnson,[3] and of control systems containing particular types of non-linear elements by other authors.[1,4,5] All these authors use this principle, or the describing-function method, as it has become called, in order to predict whether self-excited oscillations can occur, and to estimate the frequency, amplitude and stability of these oscillations if they are present. The method has also been used to obtain the response of a control system to a sinusoidal input signal.

The purpose of the paper is to review and qualify the application of the describing-function method, and also to extend its application to the evaluation of the transient response of non-linear systems.

Steady Oscillations

The system considered is the feedback control system shown in Fig. 1. In addition to a linear transfer operator, G(D) [where $D = d/dt$], which represents the low-pass filter formed by the motor and load and associated networks in the forward loop, there is also a non-linear element, whose output, $f(e)$, is an instantaneous function of the error e of the system. With the input of the system held constant it is possible for steady oscillations to occur, and this case is examined first.

Mr. Grensted is in the Department of Engineering, University of Cambridge.

The error is supposed to be oscillating sinusoidally with frequency ω and amplitude a. The output of the non-linear element will contain harmonics of this frequency, but these are supposed to be sufficiently attenuated by the low-pass filter to be

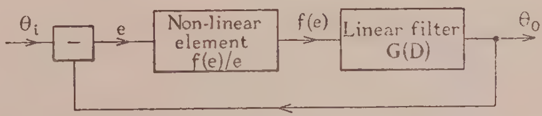


Fig. 1.—Block schematic of a non-linear feedback control system.

unimportant by the time they return round the main loop to the input of the non-linear element. The analysis therefore proceeds by neglecting all harmonics, and only the behaviour of the fundamental term throughout the system is considered.

The gain of the non-linear element is defined as the ratio of the amplitude of the fundamental component of its output to the amplitude of the sinusoidal signal applied to its input. This gain is in general complex, because a phase shift may be introduced by such non-linearities as back-lash or hysteresis. The gain of the non-linear element $f(e)$ is thus a complex function of the amplitude, a, and it may be written as $n(a)$ where

$$n(a) = p(a) + jq(a) \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$p(a) = \frac{1}{\pi a}\int_0^{2\pi} \sin\psi f(a \sin\psi)d\psi \quad . \quad . \quad . \quad (2)$$

and

$$q(a) = \frac{1}{\pi a}\int_0^{2\pi} \cos\psi f(a \sin\psi)d\psi \quad . \quad . \quad . \quad (3)$$

Thus for any given function, $f(e)$, the corresponding describing function of gain, $n(a)$, may be evaluated. Grief[4] gives a list of gain-describing functions for commonly encountered non-linearities.

For the system shown in Fig. 1 it can be readily shown that values of a and ω satisfying the equation

$$G(j\omega)n(a) + 1 = 0 \quad . \quad . \quad . \quad . \quad . \quad (4)$$

represent the amplitude and frequency of possible modes of self-excited oscillation. Eqn. (4) may be examined in the Nyquist diagram. The locus of $G(j\omega)$ with ω as a parameter and the locus of $-1/n(a)$ with a as a parameter are drawn in the complex plane. A typical example is shown in Fig. 2. Intersections of

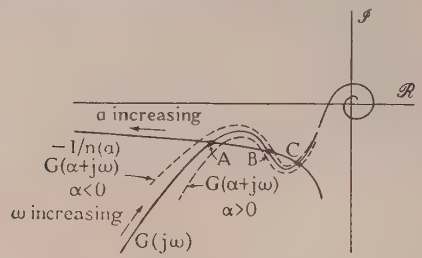


Fig. 2.—Nyquist diagram for a non-linear control system.

the two curves as A, B or C represent possible modes of steady oscillation.

The stability of these oscillations may also be determined from the diagram, but this is not so straightforward as is often supposed.

The approximation made in the analysis has been to neglect harmonics, and it has been found that this leads to errors of up to 10% in the evaluation of the amplitude and frequency of steady oscillations. In order to illustrate the type of errors that may be involved, a system was analysed in which the non-linear element was an idealized "on-off" relay, and the linear filter consisted of two exponential delays, T_1 and T_2, followed by a

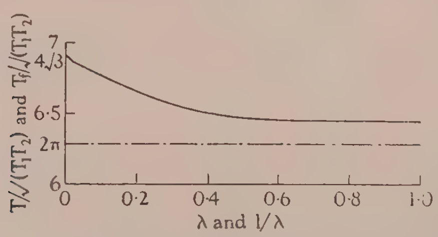


Fig. 3.—Comparison of periodic times of steady oscillations.
------ Frequency-response method: $T_f/\sqrt{(T_1T_2)}$.
——— Exact analysis: $T/\sqrt{(T_1T_2)}$.
$\lambda = \sqrt{(T_1/T_2)}$.

single integration. In Fig. 3 is shown plotted against the parameter $\sqrt{(T_1/T_2)}$ the variation of the periodic time, T_f, of steady oscillations of this system as found by the describing-function method. Also for comparison is plotted the true periodic time, T, found by an exact method.

Transient Oscillations

Transient oscillations, i.e. the behaviour of error due to a step change of input, may be considered, provided that there is only one mode of oscillation possible in the system. Otherwise, several modes of oscillation would be excited by the applied step function, and this would violate the primary assumption of the describing-function method, which is that only one component of frequency is present at any instant at the input of the non-linear element. Thus the remaining discussion is limited to a system governed by a second-order differential equation, but in this case it is possible to obtain analytic expressions for the variation with time of the frequency and damping of a transient, and hence an explicit expression for the error is derived.

It is first necessary to define frequency and damping when these are varying with time. The expression for the error, e, is

$$e = a \sin \psi \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where a is the amplitude and ψ the phase of the transient. The frequency, ω, is defined as the rate of change of phase. The damping, μ, is defined as the rate of reduction of amplitude divided by the amplitude;

thus
$$\omega = \dot{\psi}; \quad \mu = -\dot{a}/a \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Hence the expression for the error is

$$e = \varepsilon^{-\int \mu dt} \sin \int \omega dt \quad . \quad . \quad . \quad . \quad . \quad (7)$$

The system considered is that of a simple velocity-lag servomechanism in which the non-linear element has a characteristic $f(e)$. The equation for the error in this system is

$$\ddot{e} + 2c\dot{e} + f(e) = 0 \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where c is a constant.

The solution of this equation with initial conditions $\dot{e} = 0$, $e = e_0$ at $t = 0$ is the response to a step input of magnitude e_0. The describing-function method is used by substituting the assumed expression for e of eqn. (7) into eqn. (8), and neglecting all the harmonics provided by the term $f(e)$. This procedure is justified by the arguments already given in the analysis of steady oscillations. When this substitution has been made, the coefficients of the sine and cosine terms are separately equated to zero, and two important equations are derived:

$$\omega^2 = p(a) + \mu^2 - 2c\mu - \dot{\mu} \quad . \quad . \quad . \quad . \quad . \quad (9)$$

$$-\frac{\dot{a}}{a} = \mu = c + \frac{1}{2}\frac{\dot{\omega}}{\omega} + \frac{1}{2}\frac{q(a)}{\omega} \quad . \quad . \quad . \quad (10)$$

where $p(a)$ and $q(a)$ are the gain-describing functions already defined in eqns. (2) and (3).

Eqns. (9) and (10) give considerable qualitative information concerning the transient. Suppose, for instance, that $q(a) = 0$, which occurs whenever the non-linear characteristic is a single-valued function and does not introduce phase shift. Then, by eqn. (10), the damping is increased by the term $\dot{\omega}/(2\omega)$ from the value c which corresponds to the damping of the equivalent linear system. If the frequency is increasing throughout the transient then the damping is increased, and if the frequency is decreasing the damping is reduced. The contribution of the term $\dot{\omega}/(2\omega)$ to the damping can be quite significant; in an on-off control it represents one-quarter of the total damping.

The solution of eqns. (9) and (10) is possible by an iterative process. The simplest method is to use the approximate expression

$$\omega^2 \simeq p(a) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

from (9) for the value of ω in (10), which then allows (10) to be solved, giving explicit expressions for μ and a in terms of time t, and hence a more accurate value of ω may be formed from (9).

This procedure was used in the example of a control with an idealized on-off element and the expression for the error, e, was found to be

$$e = 1 \cdot 25\varepsilon^{-2t/3} \sin [173^{t/3} + 19\varepsilon^{-t/3} - 139]^\circ \quad . \quad (12)$$

The initial conditions taken where $\dot{e} = 0$, $e = 1$ at $t = 0$. In eqn. (8) $f(e) = 1$ if $e > 0$ and $f(e) = -1$ if $e < 0$; $c = 1/2$.

It is seen that the frequency of the transient increases exponentially with time but the damping remains constant. This solution, which has involved approximations in the solutions of eqns. (9) and (10) in addition to the inherent approximation of the describing-function method, is shown in Fig. 4. Peaks and zeros

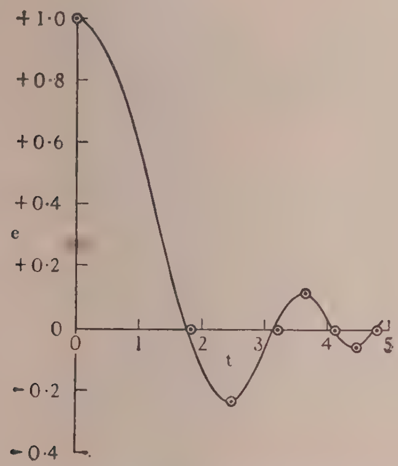


Fig. 4.—Transient solution. On-off controller with idealized relay.
○ Peaks and zeros of accurate solution.

for the exact solution of eqn. (8) are shown for comparison, and the agreement is very good.

This particular example was chosen because an exact solution could be obtained for comparison, but the method can be applied equally well to other simple non-linear systems where no comparison is possible. The main value of the method, however, lies in the fact that eqns. (9) and (10) give considerable insight into the manner in which the parameters of the system determine the nature of the transient.

References

(1) Tustin, A.: "The Effects of Backlash and of Speed-Dependent Friction on the Stability of Closed-Cycle Control Systems," *Journal I.E.E.*, 1947, **94**, Part IIA, p. 143.

(2) Kochenburger, R. J.: "A Frequency Response Method for Analysing and Synthesizing Contactor Servomechanisms," *Transactions of the American I.E.E.*, 1950, **69**, Part I, p. 270.

(3) Johnson, E. C.: "Sinusoidal Analysis of Feedback-Control Systems containing Non-Linear Elements," *ibid.*, **71**, Part II, p. 169.

(4) Grief, H. D.: "Describing Function Method of Servomechanism Analysis applied to most Commonly Encountered Non-Linearities," *ibid.*, 1953, **72**, Part II, p. 243.

(5) Nichols, B. N.: "Backlash in a Velocity Lag Servomechanism," *ibid.*, 1953, **72**, Part II, p. 462.

(6) Slemon, G. R.: "A Method of Approximate Steady-State Analysis for Non-Linear Networks," *Proceedings I.E.E.*, Paper No. 1530, September, 1953 (**100**, Part I, p. 275).

(7) Jacobsen, L. S.: "Steady Forced Vibration as influenced by Damping," *Transactions of the American Society of Mechanical Engineers, Applied Mechanics*, 1950, **53**, p. 169.

(8) Klotter, K.: "Steady State Vibrations in Systems having Arbitrary Restoring and Arbitrary Damping Forces," Proceedings of the Symposium on Non-Linear Circuit Analysis, Polytechnic Institute of Brooklyn, 1953, p. 234.

(9) Stoker, J. J.: "Non-Linear Vibrations" (Interscience Publishers, 1950).

(10) Krylov, N., and Bogoluibov, N.: "Introduction to Non-Linear Mechanics," Free translation by S. Lefschetz (Princeton University Press, 1943).

FIELD-DEPENDENT CONDUCTIVITY IN NON-UNIFORM FIELDS AND ITS RELATION TO ELECTRICAL BREAKDOWN

537.3 : 537.529 Monograph No. 128 M

PATRICK T. G. FLYNN, B.Sc.(Eng.), Ph.D., Graduate

(Digest *of a paper published in April,* 1955, *as an* Institution Monograph *and republished in September,* 1955, *in Part C of the* Proceedings.)

The calculation of the field distribution for an electrode configuration producing a non-uniform field is normally made with the assumption that the conductivity of the material between the electrodes is constant. However, it appears to be generally accepted that in insulating solids the conductivity increases when very high fields are applied. For a point-to-plane electrode system with a solid dielectric, a high field at the point will be reduced as a result of enhanced conductivity, and it therefore becomes of interest to know what the modified field distribution will be; this is of particular importance in the electrical breakdown of dielectrics in non-uniform fields.

In order to calculate the field distribution, it is necessary to find the solution of

$$\operatorname{div}(\sigma E) = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where E is the field and σ is the conductivity; the method used in the paper is to put

$$\sigma E = \sigma_0 E_t \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where σ_0 is the conductivity at some standard field. Eqn. (1) then becomes

$$\operatorname{div}(\sigma_0 E_t) = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

and since σ_0 is constant, this can be solved to find the transformed field E_t. Eqn. (2) then allows the field E to be found if the variation of σ with E is known.

Fröhlich has shown that above a critical temperature the variation of σ with E can be calculated if the energy levels in the dielectric are known; these can be obtained from measurements of the temperature dependence of the electronic conductivity in very weak fields and the temperature dependence of the field at which intrinsic breakdown occurs. A Table of the values obtained in this way for several materials is given in the paper, and Fig. 1 shows how E_t is related to E for one of them.

The method has been used to calculate the field along the axis of a point-to-plane electrode system in which the point is a hyperboloid of revolution of radius of curvature R distant x_0 from the plane, and Fig. 2 shows the result for various applied voltages V_0. Fig. 3 shows that the maximum field is considerably

---

The paper is based on Report L/T319 of the British Electrical and Allied Industries Research Association.

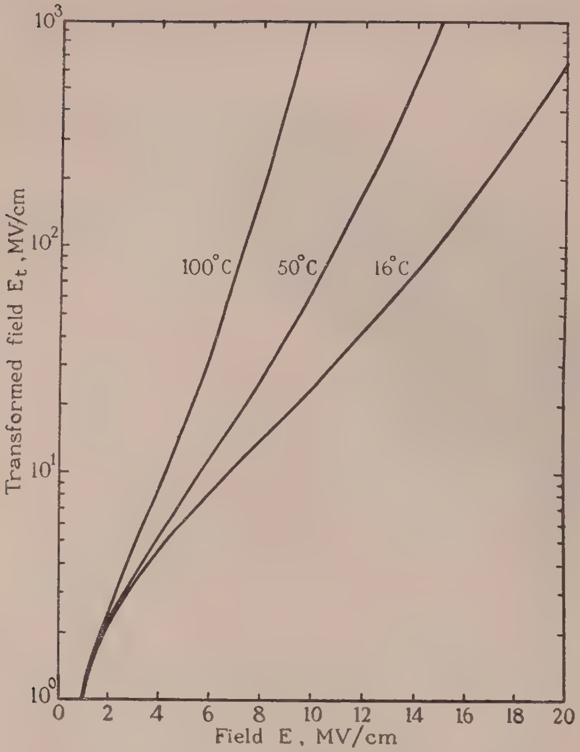


Fig. 1.—The function $E_t = (\sigma/\sigma_0)E$ for polymethylmethacrylate.

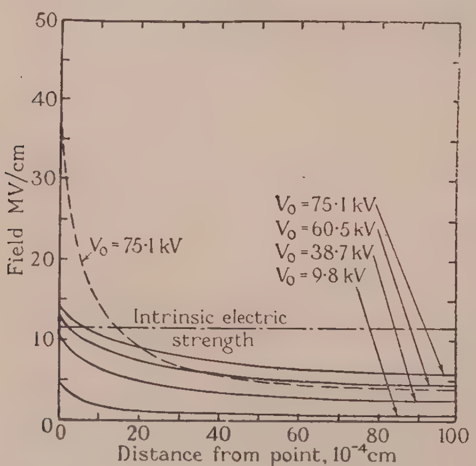


Fig. 2.—Field distribution for point with $R = 10^{-3}$ cm and $x_0 = 10^{-2}$ cm in polymethylmethacrylate at 16° C.

For comparison, dashed curve shows field calculated for constant conductivity.

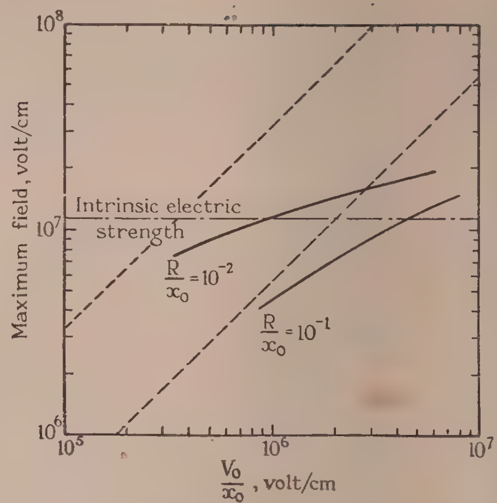


Fig. 3.—Maximum field for point-to-plane electrodes in polymethylmethacrylate at 16° C.

For comparison, dashed curves show values for constant conductivity.

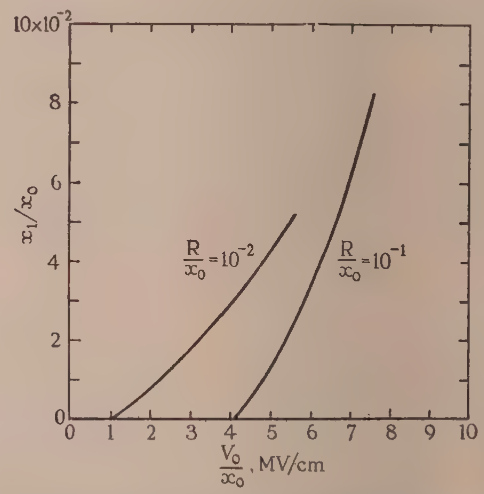


Fig. 4.—The distance x_1 over which the intrinsic electric strength is exceeded for point-to-plane electrodes in polymethylmethacrylate at 16° C.

less than that calculated for constant conductivity. This means that the voltages necessary for breakdown in a non-uniform field will be higher than those expected on the basis of a simple electrostatic calculation, and Fig. 4 shows the distance over which the intrinsic electric strength will be exceeded for one solid.

It is noted that electrons emitted by field or thermionic emission from a negative point may further modify the field and lead to polarity differences which may, however, disappear at higher temperatures.

References

(1) Eyring, C. F., Mackeown, S. S., and Millikan, R. A.: "Field Currents from Points," *Physical Review*, 1928, **31**, p. 900.

(2) Fowler, J. F., and Farmer, F. T.: "Conductivity induced in Insulating Materials by X-rays," *Nature*, 1954, **173**, p. 317.

(3) Fröhlich, H.: "Theory of Dielectric Breakdown in Amorphous Solids," *Proceedings of the Royal Society*, 1947, **A188**, p. 521.

(4) Lewis, T. J.: "The Dependence of the Dielectric Strength of Pure Liquids on Cathode Material," *Proceedings of the Physical Society*, 1953, **B66**, p. 425.

(5) Mason, J. H.: "Breakdown of Solid Dielectrics in Divergent Fields," *Proceedings I.E.E.*, Monograph No. 127 M, April, 1955 (**102** C).

(6) MAYBURG, S., and LAWRENCE, W. L.: "The Conductivity Change in Polyethylene during Gamma-Irradiation," *Journal of Applied Physics*, 1952, **23**, p. 1006.

(7) OAKES, W. G.: "The Electric Strength of Some Synthetic Polymers," *Proceedings I.E.E.*, Paper No. 780, January, 1949 (**96**, Part I, p. 37).

(8) O'DWYER, J. J.: "The Influence of Cathode Material on the Measured Breakdown Strengths of Solid and Liquid Dielectrics," *Australian Journal of Physics*, 1954, **7**, p. 400.

(9) RAMSEY, N. W.: "The Effect of Temperature on the Conductivity induced in Insulators by X-rays," *Nature*, 1953, **172**, p. 214.

(10) SIMPSON, J. H.: "Time Delay in Conduction and Breakdown Processes in Amorphous Solids," *Proceedings of the Physical Society*, 1950, **A63**, p. 86.

(11) THOMAS, A. M., and GRIFFITH, M. V.: "Intrinsic Electric Strength and Conductivity of Varnish Films and their Variation with Temperature," *Journal I.E.E.*, 1942, **89**, Part I, p. 487.

(12) TOMURA, M., and KIKUCHI, T.: "Electronic Conductivities of Amorphous Solids at High Electric Fields," *Journal of the Physical Society of Japan*, 1952, **7**, p. 538.

(13) TURNER, C. H. M., and LEWIS, W. E.: "Electrical Conductivity in Insulating Amorphous Solids," *Nature*, 1947, **159**, p. 334.

(14) VON HIPPEL, A., and MAURER, R. J.: "Electrical Breakdown of Glasses and Crystals as a Function of Temperature," *Physical Review*, 1941, **59**, p. 820.

(15) WHITEHEAD, S.: "Dielectric Breakdown of Solids" (Clarendon Press, Oxford, 1951).

BREAKDOWN OF SOLID DIELECTRICS IN DIVERGENT FIELDS

621.315.61 : 621.3.011.5 Monograph No. 127 M

J. H. MASON, B.Sc., Associate Member

(DIGEST *of a paper published in April, 1955, as an* INSTITUTION MONOGRAPH *and republished in September, 1955, in Part C of the* PROCEEDINGS.)

When insulation is subject to intense discharges, e.g. in short-time and step-by-step industrial electric strength tests, rapid breakdown occurs if the stress is sufficient to penetrate the surface and propagate through the material. It is thus of practical interest to determine the conditions necessary to propagate breakdown channels and to distinguish between the effects of stress concentration and local heating at the ends of discharge channels.

To study the effect of stress concentration alone, the propagation of breakdown channels from steel-point electrodes embedded in polythene to adjacent plane electrodes has been investigated. It is found that:

(*a*) The average divergent-field electric strength (E_{av}) and the intrinsic strength vary with temperature in a similar manner, as shown in Fig. 1. E_{av} is given by V_0/t, where V_0 is the potential of the point and t the distance between point and plane as in Fig. 2. $E_{max} \simeq E_{av}t/R(1 + 4t/R)$ is the calculated maximum stress near the point electrode.

(*b*) At 20° C the breakdown stress for a given electrode configuration is some 25% greater if the point is negative than if positive. This polarity effect decreases with increasing temperature and is negligible above 60° C.

(*c*) E_{av} decreases with increasing separation between point and plane.

(*d*) E_{av} does not vary significantly with the size of the point, unless the radius R exceeds some 10 microns (when t is between 250 and 800 microns). With larger points E_{av} increases with R such that E_{max} always exceeds the intrinsic strength, as shown in Fig. 3.

(*e*) If a surge of one polarity is followed by a surge of opposite polarity, failure occurs at much lower stress than normal for this polarity. With alternating voltage, breakdown occurs always when the point is positive.

These results indicate that stress-enhanced conductivity and space-charge accumulation reduce the stress near sharp points to values much lower than calculated geometrically (E_{max}). Breakdown channels propagate when the effective stress near the point attains the intrinsic strength of the material.

This criterion for breakdown in divergent fields explains the variation of industrial electric strength* of insulation with specimen thickness and with the electrical characteristics of the specimen and medium.

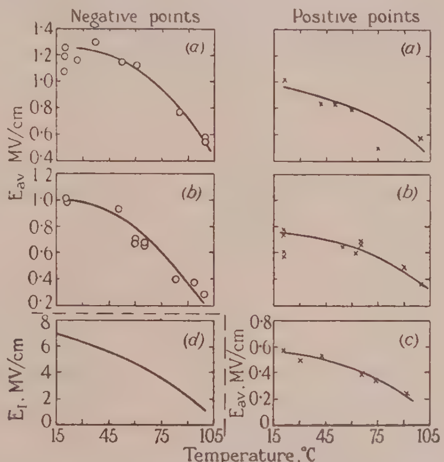


Fig. 1.—The variation of the intrinsic and average divergent-field electric strengths of polythene with temperature. Results for points of between 1·5 and 7·5 microns radius.

The electrode separation was as follows:
(*a*) 250–350 microns.
(*b*) 450–600 microns.
(*c*) 750–800 microns.
(*d*) Variation of intrinsic electric strength with temperature.

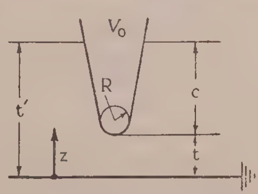


Fig. 2.—Electrode nomenclature.

The paper is based on Report Ref. L/T310 of the British Electrical and Allied Industries Research Association.

* Measured with the specimen in air or under oil, between standard electrodes such as specified by the B.S.I., A.S.T.M., etc.

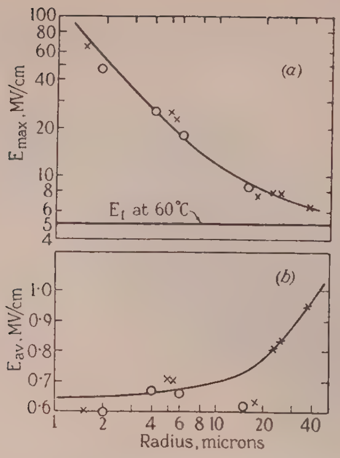


Fig. 3.—Variation of divergent-field electric strength with radius of point at 60° C.

Electrode separation, 500 microns.
× × × Positive points.
○ ○ ○ Negative points.

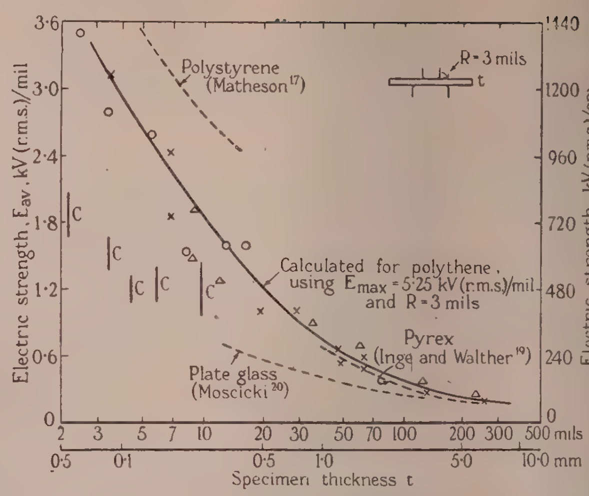


Fig. 4.—Variation with thickness of 1 min industrial electric strength of dielectric materials measured in clean oil at 20° C.

Measured values:
× × × Polythene.
△ △ △ Ebonite (Refs. 15 and 16).
○ ○ ○ Ruby mica (Ref. 14).
|C Cellulose acetate (Ref. 18).

It is assumed that discharge channels propagate through a material of thickness t when the stress E_{max} at the point of impact of a discharge of radius R attains the high temperature intrinsic strength of the material.*

The exact values of R and E_{max} are not critical, but if it were assumed that $R = 3$ mils (75 microns and $E_{max} = 5 \cdot 25$ kV (r.m.s.)/mil [3 MV(peak)/cm], the electric strength would vary with thickness, as shown by the solid curve in Fig. 4. This curve is in quite good agreement with experimental values.

The variation of electric strength with the inception voltage (V_i) of discharges in the medium (shown in Table 1) is also explained: The discharge energy, and therefore the temperature rise at the point of impact of discharges, increases proportionately to V_i^2, so that the value of E_{max} required for breakdown decreases with increase of V_i.

The variation of electric strength with time of application of alternating voltage is also considered. The long-time breakdown voltage tends asymptotically to the discharge-inception voltage V_i, but the short-time breakdown voltage decreases with increase of V_i, as shown in Table 1, owing to the increased discharge energy. Thus, if the discharge-inception voltage is increased, the a.c. working voltage is also increased, but the short-time a.c. and impulse breakdown voltage may be decreased.

TABLE 1

Factors affecting the A.C. Step-by-Step Electric Strength of Insulation. Tests with A.S.T.M. Electrodes

Specimen	Test medium					
	Air		Clean B30 transformer oil		Contaminated oil	
	V_i	Breakdown voltage	V_i	Breakdown voltage	V_i	Breakdown voltage
Polystyrene, 1 mm thick ..	3	60 after 20 sec	13·5	40 after 10 sec	6	52 when voltage increasing from 50
Polythene, 1·5 mm thick ..	2·6	50 after 25 sec	14	37·5 after 60 sec	8	47·5 when voltage increasing from 45
Cellulose acetate, 0·25 mm thick	0·9	14 after 30 sec	8·25	9 after 30 sec	7·25	11 when voltage increasing

V_i = Discharge inception voltage [kV(r.m.s.)].
Breakdown voltage [kV(r.m.s.)] determined after previous 1 min steps at lower voltage.

* Intense discharges cause local heating at the dielectric surface, so that breakdown may occur when the high-temperature intrinsic strength is exceeded irrespective of the ambient temperature.

REFERENCES

(1) MASON, J. H.: "The Deterioration and Breakdown of Dielectrics resulting from Internal Discharges," *Proceedings I.E.E.*, Paper No. 1053, January, 1951 (**98**, Part I, p. 44).
(2) MASON, J. H.: "Breakdown of Insulation by Discharges," *ibid.*, Paper No. 1471 M, March, 1953 (**100**, Part IIA, p. 149).
(3) OGILVIE, K. W.: E.R.A. Report (in course of preparation).
(4) WRIGHT, D. A.: "A Survey of Present Knowledge of Thermionic Emitters," *Proceedings I.E.E.*, Paper No. 1404 R, November, 1952 (**100**, Part III, p. 125).
(5) HUGHES, A. L., and DU BRIDGE, L. A.: "Photo-Electric Phenomena," 1932 (McGraw-Hill Book Co., New York).
(6) SALVAGE, B.: "The Dielectric Breakdown of Some Simple Organic Liquids," *Proceedings I.E.E.*, Monograph No. 2, May, 1951 (**98**, Part IV, p. 1).

(7) GOODWIN, D. W., and MCFADYEN, K. A.: "Electrical Conduction and Breakdown in Liquid Dielectrics," *Proceedings of the Physical Society*, 1953, **66**, p. 85.

(8) CALDERWOOD, J. H., COOPER, R., and WALLACE, A. A.: "An Experimental Investigation of Factors influencing the Electric Strength of Potassium Chloride Crystals," *Proceedings I.E.E.*, Paper No. 1457 M, March, 1953 (**100**, Part IIA, p. 105).

(8A) COOPER, R., and GROSSART, D. T.: "The Influence of Cathode Material on the Electric Strength of Potassium Bromide Crystals," *Proceedings of the Physical Society*, 1953, **66** B, p. 716.

(9) HAWLEY, E. M.: "The Variation with Temperature of the Electric Strength of Polyisobutylene subjected to Discharges," E.R.A. Report Ref. L/T290, 1953.

(10) WHITEHEAD, S.: "Dielectric Breakdown of Solids" (Clarendon Press, 1951), Chapter 1.

(11) WHITEHEAD, S.: "Breakdown of Solid Dielectrics" (Benn, 1932), Chapter 3.

(12) LIAO, T. W., and ANDERSON, J. G.: "Propagation Mechanism of Impulse Corona and Breakdown in Oil," *Communication and Electronics*, 1953, No. 9, p. 641.

(13) MOSCICKI: *Electrotechnische Zeitschrift*, 1904, **25**, p. 527.

(14) LEWIS, A. B., HALL, E. L., and CALDWELL, F. R.: "Some Electrical Properties of Micas," *Journal of Research of the National Bureau of Standards*, 1931, **7**, p. 403.

(15) FARMER, F. M.: "Dielectric Strength of Thin Insulating Materials," *Transactions of the American I.E.E.*, 1913, **32**, p. 2097.

(16) PEEK, F. W.: "Dielectric Phenomena in High Voltage Engineering" (New York, 1916).

(17) MATHESON, L. A., and GOGGIN, W. C.: "Electrical Application of Polystyrene," *Industrial and Engineering Chemistry*, 1939, **31**, p. 334.

(18) "Cellulose Derivative Films for Insulating Purposes," E.R.A. Report Ref. A/T118, 1950.

(19) INGE, L., and WALTHER, A.: *Archiv für Electrotechnik*, 1930, **24**, p. 88.

(20) STANDRING, W. G., and CHAPPEL, D. W. G.: "The Breakdown of Bitumen under Impulse Voltages," E.R.A. Report Ref. L/T113, 1943.

(21) INGE, L., and WALTHER, A.: "Elektrische Entladungen Kristallen," *Zeitschrift für Physik*, 1930, **64**, p. 830.

(22) HOWARD, P. R.: "Impulse Puncture Characteristics of Mass-Impregnated Paper-Insulated Cables," *Proceedings I.E.E.*, Paper No. 1515 S, June, 1953 (**100**, Part II, p. 315).

(23) DAKIN, T. W., and WORKS, C. N.: "Impulse Dielectric Strength Characteristics of Liquid Impregnated Pressboard," *Transactions of the American I.E.E.*, **71**, Part I, p. 321.

(24) FLYNN, P.: "Field Dependent Conductivity in Non-Uniform Field Distributions and its Relation to Electrical Breakdown," *Proceedings I.E.E.*, Monograph No. 128 M, April, 1955 (**102** C).

(25) HUNTER, E., and OAKES, W. G.: "The Effect of Temperature on the Density of Polythene," *Transactions of the Faraday Society*, 1945, **41**, p. 49.

THE GRAPHS OF ACTIVE NETWORKS

621.372.4/.5 : 621.3.012 Monograph No. 129 R

W. S. PERCIVAL, B.Sc., Associate Member

(DIGEST *of a paper published in April*, 1954, *as an* INSTITUTION MONOGRAPH *and republished in September*, 1955, *in Part C of the* PROCEEDINGS.

In a previous paper, it was shown that the graph of a passive network, which does not contain transformers, could be expanded as the sum of a number of simpler graphs. From these graphs the set of trees, which is equal to the determinant of the network, could be written down by inspection.

In a further paper,[2] which will be referred to as "Determinants," it was shown that 3- or 4-terminal elements, such as valves and transformers, could be represented as graphs in terms of an element $h_{ij.km}$ comprising four nodes and two directed branches. This makes it possible to set up the determinant of any network directly from the graph of that network.

The present paper lays a basis for a theory of graphs of networks containing 3- or 4-terminal elements.

Fig. 1 shows the graphs of the various network elements as described in "Determinants": (*a*) shows the fundamental element $h_{ij.km}$ which is such that a voltage v_{jm} between the nodes j and m, with m positive, produces a current i_{ik} from k to i equal to $h_{ij.km}v_{jm}$; (*b*) shows an ordinary 2-terminal element obtained from $h_{ij.km}$ by identifying i and j and also m and k; (*c*) shows a valve; (*d*) shows a pair of identical valves in push-pull; (*e*) shows a symmetrical two-way admittance, which may be that of a transformer; $f_{kj.im}$, shown in (*f*), is a fictitious element equal to the transfer admittance which it is required to evaluate.

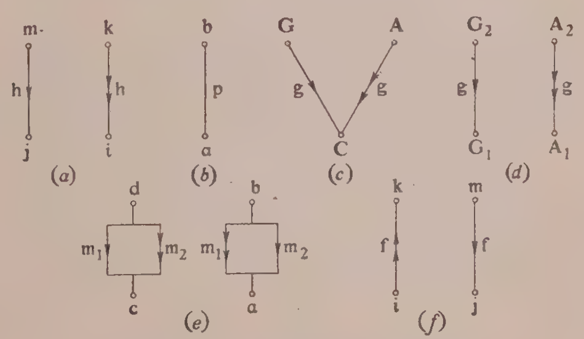


Fig. 1.—Graphs of the network elements.

(*a*) $h_{ij.km}$
(*b*) $P_{aa.bb}$
(*c*) $g_{CC.AG}$
(*d*) $g_{A1G1.A2G2}$
(*e*) $(m_{ac.bd}) + (m_{ca.db})$
(*f*) $f_{kj.im} = Y_{ij.km}$

Eqn. (1) gives the H-matrix (as defined in "Determinants") of

Mr. Percival is with Electric and Musical Industries, Ltd.

the element $h_{ij.km}$, from which the H-matrices of all other elements can be deduced:

$$(h_{ij.km}) = \begin{array}{c|cc} & j & m \\ i & h & -h \\ k & -h & h \end{array} \quad . \quad . \quad . \quad . \quad (1)$$

The H-matrix of a network is the sum of the H-matrices of its elements, and the ordinary D-determinant is equal to any first cofactor of the determinant of the H-matrix of the network.

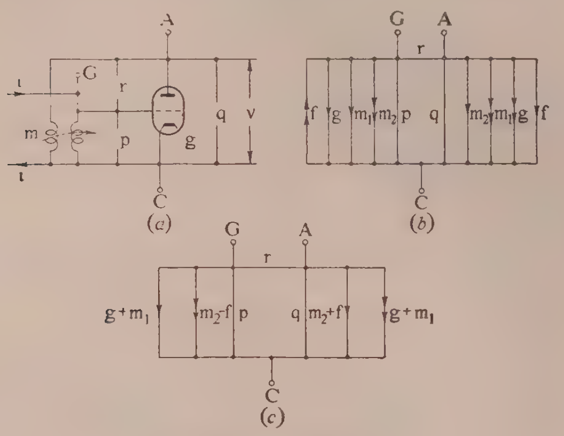


Fig. 2.—Triode network.

Fig. 2(a) is the conventional representation of a network fed with a current i and from which it is required to obtain the output voltage v. Fig. 2(b) shows the graph of this network in which $f_{GC.CA} = Y_{CC.GA} = i/v$. In Fig. 2($c$) elements in parallel are replaced by a suitably labelled single element.

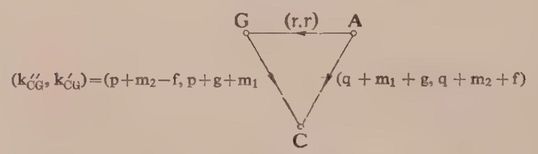


Fig. 3.—The reduced graph.

Fig. 3 shows the "reduced graph," in which branches in parallel are replaced by a single branch and suitably labelled. Each of these branches represents a current branch, denoted by the first term in the brackets, in parallel with a voltage branch, denoted by the second term in the brackets. Hence the reduced graph can be replaced by two graphs, a current graph as in

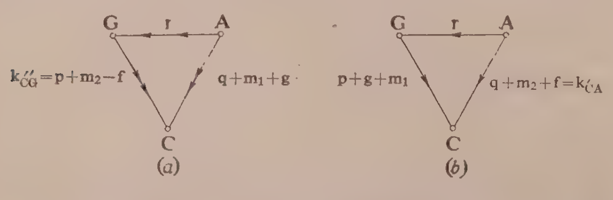


Fig. 4.—Current and voltage graphs.

Fig. 4(a) composed of current branches only, and a voltage graph as in Fig. 4(b) composed of voltage branches only.

The current graph must be in accordance with Kirchhoff's law for currents, while the voltage graph must be in accordance with Kirchhoff's law for voltages. The two graphs are linked by the admittances of the elements. The nodes of the current graph correspond to the rows of the H-matrix, while the nodes on the voltage graph correspond to the columns. To every operation on the H-matrix there corresponds an operation on the graphs, and this correspondence extends even to operations in which the nodes on the two graphs are different.

Each matrix element of the H-matrix can be obtained from Fig. 3 by the rule that h_{ij} is equal to the sum of the network elements the current branches of which terminate on node i and the voltage branches on node j, the sign being positive if the arrows on both branches point either towards or away from the nodes i and j, and negative otherwise. If the fictitious element f is included, $D = 0$. The transfer admittance is then equal to the ratio of $D\,(f = 0)$ to the coefficient of f in D.

The expression k''_{ik}, which has been used as a label for a current branch in the current graph, is termed the current-branch admittance function; the expression k'_{jm}, which has been used as a label for a voltage branch in the voltage graph, is termed the voltage-branch admittance function; while the sum of the elements common to k''_{ik} and k'_{jm} is termed the branch-pair admittance and is written $k_{ij.km}$. An algebra of the graphs is developed which includes relations between these expressions and the matrix elements h_{ij}.

It was shown in "Determinants" that the derivative of an H-determinant with respect to the element $h_{ij.km}$ is $(i + k, j + m)H$, where $(i + k, j + m)$ is an addition operator which deletes row k, replaces row i by the sum of the rows i and k, performs the corresponding operations on the columns j and m, and affixes the sign $(-1)^{k+m}$. It is shown that the corresponding operation on the current and voltage graphs involves short-circuiting the nodes i and k on the current graph and also the

Fig. 5.—$\partial H/\partial g$ = Derivative of the graphs of Fig. 4 with respect to $g_{CC.AG}$.

nodes j and m on the voltage graph. Fig. 5 shows the derivative of the graph of Fig. 4 with respect to the valve element g.

On a graph, all the elements of which are ordinary elements of the form $h_{il.jj}$, a tree is defined as a set of branches which connect all the nodes but enclose no meshes. On such a graph a tree can be formed by a set of operations of the form $h_{ii.jj}\partial/\partial h_{ii.jj}$. On the current and voltage graphs a set of the more general operations $h_{ij.km}\partial/\partial_{ij.km}$ gives rise to a pair of trees, one on each graph. A tree on the graphs of a network is therefore defined as a set of network elements such that the current branches form a tree on the current graph and the voltage branches form a tree on the voltage graph.

With this extended definition of a tree, and with an appropriate rule of signs, the set of trees T on any network is equal to the determinant D of the network. Thus the usual theorem is extended to include networks containing valves and transformers. Moreover, if the element f is included, the equation $T = 0$ gives the value of the transfer admittance. An example is given embodying Wheeler's concept[3] of the ideal transformer-repeater.

To solve simple networks, consisting of 2-terminal elements only, the choice of method is of little importance. If the network is more complicated the H-determinant method is generally quicker. If the labour of evaluating the determinant is still considerable, the tree method is more rapid since there are no terms to be cancelled.

For networks containing valves and transformers it has been found of considerable advantage to construct the reduced graph of the network. Either the *D*-determinant or the *H*-determinant can then be set up from the graph. If valves or transformers are present the tree method is inevitably more complicated than for networks containing 2-terminal elements only. However, the labour of expanding a very large determinant can be considerably reduced by expanding the graphs as the sum of simpler graphs and then adding the determinants of these graphs.

REFERENCES

(1) PERCIVAL, W. S.: "The Solution of Passive Electrical Networks by Means of Mathematical Trees," *Proceedings I.E.E.*, Paper No. 1492 R, May, 1953 (**100**, Part III, p. 143).
(2) PERCIVAL, W. S.: "Improved Matrix and Determinant Methods for Solving Networks," *ibid.*, Monograph No. 96 R, April, 1954 (**101**, Part IV, p. 258).
(3) WHEELER, H. A.: Wheeler Monographs, No. 5, Vol. 1, Wheeler Laboratories.

RESIDUAL EFFECTS IN THE CAMPBELL BRIDGE METHOD FOR THE ABSOLUTE MEASUREMENT OF RESISTANCE, AND A NOTE ON A NEW BRIDGE

621.317.331 Monograph No. 130 M

Professor N. F. ASTBURY, M.A., Sc.D., F.Inst.P., Member

(DIGEST *of a paper published in April, 1955, as an* INSTITUTION MONOGRAPH *and republished in September, 1955, in Part C of the* PROCEEDINGS.)

The technical problem in the absolute determination of resistance is the comparison of the measure of a physical model of the unit of resistance with the quotient of inductance and time, the former being referred to the standard of length and the permeability of free space, and the latter to the mean solar second. Both generator[2] and circuit methods[3–6] have been used in absolute determinations. Amongst the latter is that due to Albert Campbell,[1,6] which is one of the few methods in which pure alternating current, and not commutated current, is used.

This method was developed at the National Physical Laboratory[6] in 1933–36, and a new determination[1] was made in 1951–54. The circuit is shown, in its essentials, in Fig. 1. The main balance condition at a frequency $\omega/2\pi$ c/s is, ideally, $\omega^2 M_1 M_2 = rR$. If the ratio of r/R be λ, the value of r in absolute measure is $\omega\sqrt{(M_1 M_2 \lambda)}$. The balance conditions can be established for eight combinations of the connections to the windings of M_1 and M_2 (M_1 and M_2 being always of opposite sign), and when allowance has been made for residual effects in the resistors and inductors, the eight combinations should all give the same result. In both determinations, however, it was found that, although the mean of the results of the eight combinations was reproducible to within a few parts per million, there were variations within the groups of several parts in 10^5. It was realized that it would be necessary to correct for coupling between supply and detector loops, and an attempt was made to do this experimentally with a controlled eddy-current circuit. Nevertheless the systematic differences persisted.

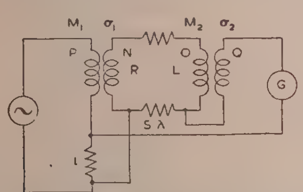


Fig. 1.—N.P.L. Campbell bridge.

It is shown in the present paper that these differences can be explained if it is assumed that the method of correction noted above was ineffective, and the circuit conditions are re-evaluated on the basis of the existence of an eddy-current link (due in practice to connectors, switches, etc.) between supply and detector circuits. There then appear to be four distinct results reproduced twice within a group of eight determinations: the mean of these will be nearly, but not exactly, the "true value," i.e. many of the disturbing terms will cancel out, the residual uncertainty being probably of the order of one or two parts per million. The original results[6] seem to support this. The eight results in any one group appear to be a duplicate set of four differentiated by random errors.

The analysis has been checked experimentally on a new bridge circuit shown in Fig. 2. While this circuit possibly has some advantages in regard to earthing and the definition of common-point connections for the inductors, it has the same essential features as the Campbell circuit, and eddy-current coupling between supply and detector loops will produce similar systematic variations when the coil connections are interchanged. Here, there are only four possible combinations.

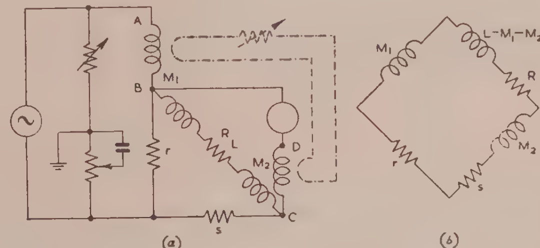


Fig. 2.—New bridge and transformation.

A model bridge, designed to work at 1 000 c/s, was built using toroidal inductors and containing a minimum of metal, so that adventitious eddy-current coupling was reduced to negligible levels. The effects of eddy-current coupling were then simulated

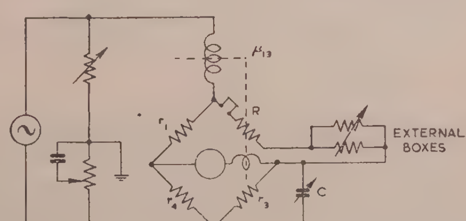


Fig. 3.—Auxiliary bridge for μ_{13}: tertiary circuit coupling indicated by dotted line.

Professor Astbury is at the University College of Khartoum.

and controlled by means of a tertiary circuit, indicated by dotted lines in Fig. 2. Quantitative assessment was made in terms of the effective mutual impedance μ_{13} of the supply and detector loops, which was measured on an auxiliary bridge (Fig. 3). In

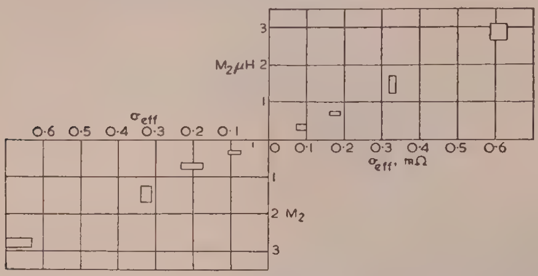


Fig. 4.—Correlation of M_2 readings at balance with σ_{eff}.

Upper-quadrant coil connections

Primary		Secondary	
M_1	M_2	M_1	M_2
+	+	+	+
−	−	−	−

Lower-quadrant coil connections

Primary		Secondary	
M_1	M_2	M_1	M_2
−	+	−	+
+	−	+	−

Fig. 4 the change, ΔM, in the balancing value of M_2, caused by coil reversals, is plotted against the resistive component, σ_{13} of μ_{13}, the scatter of the observations being indicated. These results, and a plot of ΔM and σ_{13} against the conductance of the tertiary circuit, appear to establish the validity of the hypothesis of the origin of the systematic errors.

It appears that the new bridge could be used in such a way as to make a correct allowance for eddy-current coupling. Referring to Fig. 2, if measurements are made of the mutual impedances $\sigma_1 + j\omega M_1$ for the system ABC, $\sigma_2 + j\omega M_2$ for BCD and $\sigma_3 + j\omega M_3$ for ABCD, the main equation of balance can be put into the form

$$\omega^2 M_1 M_2 (1 + \alpha) = rR(1 + \beta)$$

where α and β are small terms involving M_3 and σ_3, respectively, Thus the effects of eddy-current coupling would be included with the other residuals, and no external compensation would be needed.

References

(1) Rayner, G. H.: "An Absolute Measurement of Resistance by Albert Campbell's Bridge Method," *Proceedings I.E.E.*, Monograph No. 95 M, April, 1954 (**101,** Part IV, p. 251).
(2) Vigoureux, P.: *Collected Research Reports of the National Physical Laboratory*, London, **24**, Paper No. 8.
(3) Grüneisen and Giebe: *Annalen der Physik* (*Leipzig*), 1920, **368**, p. 179.
(4) Curtis, H. L., Moon, C., and Sparks, C. M.: *Journal of Research of the National Bureau of Standards*, 1936, **16**, p. 1.
(5) Thomas, J. L., Peterson, C., Cooter, I. L., and Kotter, F. R.: *ibid.*, 1949, **43**, p. 291.
(6) Hartshorn, L., and Astbury, N. F.: *Philosophical Transactions of the Royal Society*, A, 1937, **236**, p. 423.
(7) Jouaust, R., Picard, M., and Hérou, R.: *Bulletin de la Société Française des Electriciens*, 1938.
(8) Butterworth, S.: *Proceedings of the Physical Society*, 1921, **33**, p. 312.

THE THEORY AND DESIGN OF COAXIAL RESISTOR MOUNTS FOR THE FREQUENCY BAND 0–4000 Mc/s

621.316.89.029.6 Monograph No. 132 R

I. A. HARRIS, Associate Member.

(Digest *of a paper published in May*, 1955, *as an* Institution Monograph *and to be republished in Part C of the* Proceedings.)

Some methods of realizing resistance without reactance, constant from d.c. to ultra-high frequencies, employ a cylindrical resistive film as the inner conductor of a short, uniform coaxial line.[1,2] With this design, the size of the resistor and the power rating are restricted by the upper frequency limit. The basis of another method is shown in Fig. 1, in which the characteristic impedance

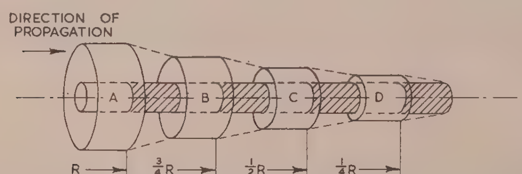


Fig. 1.—Cylindrical film resistor divided into four equal parts, with sections of loss-free coaxial line interposed.

related solely to the diameter ratio at any section is equal to the resistance between that section and the termination. On increasing the number of divisions indefinitely and applying the usual formula for lossless coaxial lines, there results the exponential profile:[3]

$$r = a \exp (2\pi R_0 w/\zeta) \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where $\zeta = (\mu/\epsilon)^{\frac{1}{2}}$, μ and ϵ being the absolute permeability and permittivity respectively of the coaxial space; R_0 is the resistance per unit length, w is distance from the termination, a is the resistor radius, and r the radius of the outer.

This design, however, is imperfect in so far as the wavefronts in Fig. 1 are assumed to be planar, whereas they are in fact curved surfaces because the electric field must meet the outer conductor at right angles and the resistor at an angle depending on R_0 and a. To take account of the curved wavefronts, the design is better approached by taking a number of conical line sections, rather than the cylindrical line sections[7] of Fig. 1. The resulting profile of the outer conductor is a tractrix, and an approximate formula for r as a function of w is

$$r(w) = \frac{a(1 + s^2/4) \exp [(w + \Delta w)/\rho_0]}{1 + (s^2/4) \exp [2(w + \Delta w)/\rho_0]} \quad . \quad . \quad (2)$$

Mr. Harris is in the Laboratories of the Aeronautical Inspection Directorate, Ministry of Supply.

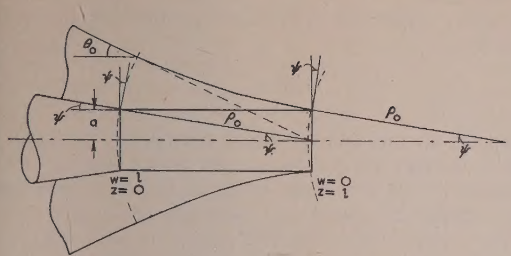


Fig. 2.—Nomenclature and geometrical arrangement for resistor mount with coaxial cone input terminals.

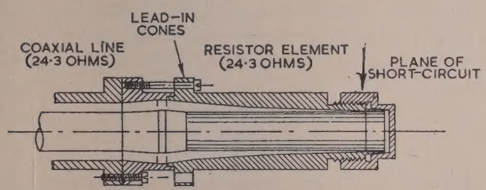


Fig. 3.—Section of experimental "terminal" resistor mount, of 24·3 ohms.

in which $\Delta w = (as/2)\,[\exp\,(2w/\rho_0) - 1]$, $s = \sin\psi = 2\pi a R_0/\zeta$ and $\rho_0 = a/s$ (see Fig. 2).

To allow for the field which penetrates the film into the enclosure made by it, formula (2) is multiplied by the factor $(1 + \epsilon_r s^2/4)$, where ϵ_r is the relative permittivity of the enclosure compared with that of the coaxial space.

Experimental results with the mount shown in Fig. 3 show variations in resistance within 1% from direct current to the highest frequency measured (3 450 Mc/s), with a phase angle within 3° at the highest frequency.

References

(1) Crosby, D. R., and Pennypacker, C. H.: "Radio Frequency Resistors as Uniform Transmission Lines," *Proceedings of the Institute of Radio Engineers*, 1946, **34**, p. 62.

(2) Kohn, C. T.: "The Design of a Radio Frequency Coaxial Resistor," *Proceedings I.E.E.*, Monograph No. 83, November, 1953 (**101**, Part IV, p. 146).

(3) Montgomery, C. C., *et al.*: "Technique of Microwave Measurements," M.I.T. Radiation Laboratory Series No. 11, p. 725 (McGraw-Hill, New York, 1947).

(4) Heaviside, O.: "Electromagnetic Theory" (Spon, London, 1951), Vol. 1, p. 386.

(5) Bromwich, T. J. I'A.: "Electromagnetic Waves," *Philosophical Magazine*, 1919, **38**, p. 143.

(6) See, for example, Piaggio, H. T. H.: "Differential Equations" (Bell, London, 1933), p. 109.

(7) British Patent Specification No. 670339.

(8) McLachlan, N. W.: "Bessel Functions for Engineers" (University Press, Oxford, 1941).

THE CORRELATION BETWEEN DECAY TIME AND AMPLITUDE RESPONSE

621.372.4/.5 : 621.3.011 Monograph No. 139 R

S. DEMCZYNSKI, Dipl.Eng., Graduate.

(Digest *of a paper published in July,* 1955, *as an* Institution Monograph *and to be republished in Part C of the* Proceedings.)

The aim of the present work is to investigate the correlation between decay time and delay time of the indicial response of certain types of four-terminal networks on the one hand, and, on the other, certain easily measurable parameters of their normalized amplitude response, namely the ratio f_3/f_6 and the peak value H_1 (see Fig. 1). In order to obtain more uniform results and a better basis for comparison of various networks, the transients considered are all responses to the unit voltage step. Such transients are called *indicial responses*.

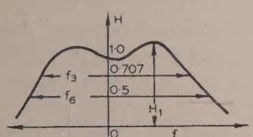


Fig. 1.—Steady-state amplitude characteristic.

The decay time t_d (see Fig. 2) is defined as the time taken from the mid-point amplitude of the indicial response to the moment when the envelope of the indicial response deviates from the steady-state by a prescribed proportion of the amplitude, e.g. $t_d(1\%)$ or $t_d(0{\cdot}1\%)$. The delay time t_l is defined as the time taken by the indicial response curve to reach the half-amplitude point.

The circuits considered are all minimum-phase passive networks consisting of lumped parameters and comprising some low-pass and band-pass filters. Some formulae giving the relations between the parameters of amplitude response on the one hand and the decay time of the indicial response on the other are deduced by comparing directly the expressions representing these two states for the given type of network. Others are found by interpolation of the results obtained for many types of circuit.

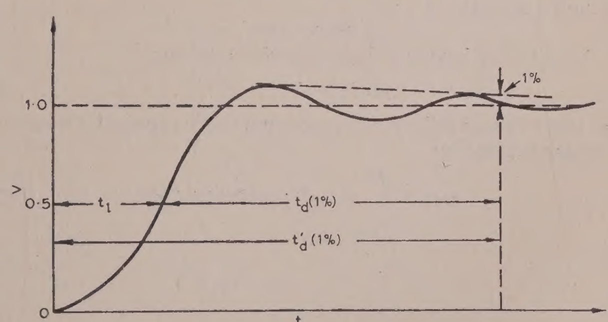


Fig. 2.—Indicial response.

The former present the existing relations more exactly; the latter are more approximate but at the same time simpler and more suitable for the designer to use.

Mr. Demczynski is with Marconi's Wireless Telegraph Co., Ltd.

Decay Time

Increasing the slope of the amplitude response results in an increase of the overshoot, which may be considered as the initial amplitude of a quasi-sinusoidal exponentially-damped oscillation on the "top" of the indicial response. It also reduces the attenuation of this oscillation. Both effects increase the decay time. However, when there is no oscillation, as in the case of n RC stages, the decay time, as defined here, is reduced with increasing slope. The apparent connection between the decay time and the slope of the amplitude response suggests the possibility of finding a general functional relationship between these two quantities.

To find this relationship, five different types of networks are selected and the decay times for various values of Q and n are calculated. At the same time the corresponding values of $\delta = H_1 f_2/f_6$ are found. The resultant curves are presented in the form $t_d(1\%)f_6 = \gamma = \mathrm{F}(\delta)$. By interpolating all the curves by a single straight line, on the principle of least total fractional square error, the statistical formulae, valid for all circuits considered, are found.

Because of the variety of circuits taken into consideration, these approximate formulae can be expected to be valid for many other passive networks and may well serve as a first, rough indication for the designer. It is found that, in order to decrease decay time, the slope of the amplitude response must be decreased. For the same slope, a circuit having an amplitude response with peaks will have a longer decay time than one with a flat response.

Another approach to the problem is based on the fact that any passive minimum-phase network is completely defined by the location of its poles and zeros in the p-plane, except for a possible constant multiplier. Hence certain classes of circuit can be distinguished, e.g. filters having frequency response with one, two, or more single or multiple poles. An attempt is made here to derive the general formulae, valid for all circuits belonging to a given class, which give the functional dependence in the form: $t'_d(1\%)f_6 = \gamma'(1\%) = \mathrm{F}(f_3/f_6)$, where $t'_d(1\%)$ is the total decay time, defined as the time taken from the moment of application of input unit voltage step to the moment when the envelope of the transient response deviates from the steady state by a prescribed proportion of the amplitude, e.g. $t'_d(1\%)$.

The relations of the above type are derived for circuits with one single pole, two single poles, one multiple pole, for the staggered circuits and n stages of critically coupled and over-coupled circuits.

Delay Time

If to a filter having a frequency characteristic

$$H(f) = H(f)\varepsilon^{j\phi(f)} \quad \ldots \ldots \quad (1)$$

an input voltage pulse having spectrum $G(f)$ is applied, the output voltage is given by

$$v(t) = \int_{-\alpha}^{\alpha} G(f)H(f)\varepsilon^{j[2\pi ft+\phi(f)]}df \quad \ldots \quad (2)$$

Hence every frequency component of the output voltage is delayed by

$$\Delta t = \frac{\phi(f)}{2\pi f} \quad \ldots \ldots \quad (3)$$

with respect to the corresponding component of input pulse. Hence, in contrast to the decay time, which is more obviously connected with the amplitude response, the delay time seems to be more readily correlated with the phase response.

From the character of the spectrum of the unit voltage step it is evident that the bulk of the energy is transmitted by the components near to zero frequency. This effect is increased by the discrimination introduced by the filter itself. In the frequency range occupied by these components, the phase response of the filter is usually almost linear; hence their delay time, which may be regarded approximately as the delay time of the whole pulse, is given by

$$t'_l = \left(\frac{d\phi}{d\omega}\right)_{\omega=0} \quad \ldots \ldots \quad (4)$$

This equation has been checked for all circuits previously considered by comparing the results it gives with the exact values of t.

It was found that eqn. (4) gives a fairly sufficient degree of approximation for multiple-stage circuits, but it seems to be quite wrong for the single-stage circuits here considered.

This may be attributed to the fact that, in the single-stage circuit the harmonics lying outside the linear range of phase response are rather poorly attenuated, and the previously explained idea of delay time has no meaning so that eqn. (4) has no justification in this case.

References

(1) Guillemin, E. A.: "Communication Networks," Vol. II (Wiley, 1945).

(2) Bode, H.: "Network Analysis and Feedback Amplifier Design" (Van Nostrand, 1945).

(3) Waliman, H.: "Vacuum Tube Amplifiers," M.I.T. Radiation Laboratory Series Vol. 17 (McGraw-Hill, 1947).

(4) Carson, S., and Jaeger, J. C.: "Operational Methods in Applied Mathematics" (Oxford University Press, 1941).

(5) Gardner, M., and Barnes, J.: "Transients in Linear Systems" (Wiley, 1942).

(6) Levy, M.: "The Impulse Response of Electrical Networks," *Journal I.E.E.*, 1943, **90**, Part III, p. 153.

(7) Jelonek, Z.: "Transient Response," Lectures (unpublished).

(8) Kaliman, H. E., and Spencer, R. E.: "Transient Response," *Proceedings of the Institute of Radio Engineers*, 1945, **31**, p. 169.

(9) Jaworski, Z.: Diploma Thesis, 1950, Polish University College.

PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

Part B. RADIO AND ELECTRONIC ENGINEERING (INCLUDING COMMUNICATION ENGINEERING), SEPTEMBER 1955

CONTENTS

	PAGE
The Mechanism of Sub-Harmonic Generation in a Feedback System . . . J. C. WEST, B.Sc., Ph.D., and J. L. DOUCE, M.Sc., Ph.D.	569
The Response of Remote-Position-Control Systems with Hard-Spring Non-Linear Characteristics to Step-Function and Random Inputs. J. C. WEST, B.Sc., Ph.D., and P. N. NIKIFORUK, B.Sc., Ph.D.	575
Discussion on the above two Papers	594
The Adjustment of Control Systems for Quick Transient Response . . . A. T. FULLER, M.A.	596
The Integral-of-Error-Squared Criterion for Servo Mechanisms . . . H. H. ROSENBROCK, B.Sc.(Eng.), Ph.D.	602
Discussion on "The Manchester–Kirk o'Shotts Television Radio-Relay System"	607
High-Speed Electronic-Analogue Computing Techniques . . . D. M. MACKAY, B.Sc., Ph.D.	609
Artificial Reverberation . . . P. E. AXON, O.B.E., M.Sc., Ph.D., C. L. S. GILFORD, M.Sc., and D. E. L. SHORTER, B.Sc.(Eng.)	624
The Subjective Discrimination of Pitch and Amplitude Fluctuations in Recording Systems. A. STOTT, M.A., and P. E. AXON, O.B.E., Ph.D., M.Sc.	643
A Study of the Long-Term Emission Behaviour of an Oxide-Cathode Valve . . . G. H. METSON, M.C., Ph.D., M.Sc., B.Sc.(Eng.)	657
Some Experiments on the Breakdown of Heater-Cathode Insulation in Oxide-Cathode Receiving Valves. G. H. METSON, M.C., Ph.D., M.Sc., B.Sc.(Eng.), E. F. RICKARD, B.Sc., and F. M. HEWLETT, B.Sc.	678
Discussion on "Noise Generation in Crystals and Ceramic Forms of Barium Titanate when subjected to Electric Stress"	683
A Milliwattmeter for Centimetre Wavelengths . . . A. C. GORDON-SMITH	685
V.H.F. Aerials for Television Broadcasting . . . G. J. PHILLIPS, M.A., Ph.D., B.Sc.	687
On Distributed Amplification . . . D. G. SARMA, M.Sc.	689
Theory of Imperfect Waveguides: The Effect of Wall Impedance . . . A. E. KARBOWIAK, Ph.D.	698
Discussion on "A General Experimental Method to determine the Properties of Artificial Media at Centimetre Wavelengths, applied to an Array of Parallel Metallic Plates"	708
Measurements of the effect of Rain, Snow and Fogs on 8·6 mm Radar Echoes . . . N. P. ROBINSON, M.A.	709
Discussion on "Submerged Telephone Repeaters for Shallow Water" and "The British Post Office Standard Submerged-Repeater System for Shallow-Water Cables, with special mention of the England–Netherlands System"	714
Discussion on "A Short Modern Review of Fundamental Electromagnetic Theory"	716
Discussion on "A Transatlantic Telephone Cable"	717
Digest of a Paper: Current Summations with Current Transformers . . . A. HOBSON, M.Sc.Tech.	719
Digests of Institution Monographs	721

Declaration on Fair Copying.—Within the terms of the Royal Society's Declaration on Fair Copying, to which The Institution subscribes, material may be copied from issues of the *Proceedings* (prior to 1949, the *Journal*) *which are out of print and from which reprints are not available.* The terms of the Declaration and particulars of a Photoprint Service afforded by the Science Museum Library, London, are published in the *Journal* from time to time.

Bibliographical References.—It is requested that bibliographical reference to an Institution paper should always include the serial number of the paper and the month and year of publication, which will be found at the top right-hand corner of the first page of the paper. This information should precede the reference to the Volume and Part. *Example.*—SMITH, J.: "Reflections from the Ionosphere," *Proceedings I.E.E.*, Paper No. 3001 S, December, 1954 (102 B, p. 1234).

THE BENEVOLENT FUND

Since the foundation of the **Benevolent Fund** over **£100 000** has been distributed in grants to members or their dependants who have been disabled by illness or have found themselves in straitened circumstances.

In the number of subscribers, the Fund compares very favourably with those of similar organizations, but there are still many members of The Institution who could help, and even a small amount per head would greatly strengthen the Fund.

Subscriptions and Donations may be sent by post to

THE HON. SECRETARY, THE INCORPORATED BENEVOLENT FUND OF THE INSTITUTION OF ELECTRICAL ENGINEERS, SAVOY PLACE, W.C.2

or may be handed to one of the Local Honorary Treasurers of the Fund.

☆

THE FUND HELPS MEMBERS OF THE INSTITUTION AND THEIR DEPENDANTS IN TIMES OF ILL-HEALTH OR OTHER DIFFICULTIES

☆

Local Hon. Treasurers of the Fund:

EAST MIDLAND CENTRE . . . . *R. C. Woods*
IRISH BRANCH . . . . *A. Harkin, M.E.*
MERSEY AND NORTH WALES CENTRE . *D. A. Picken*
NORTH-EASTERN CENTRE . . . . *D. R. Parsons*
NORTH MIDLAND CENTRE . . . . *J. G. Craven*
SHEFFIELD SUB-CENTRE . . . .
NORTH-WESTERN CENTRE . . . . *W. E. Swale*
NORTH LANCASHIRE SUB-CENTRE *G. K. Alston, B.Sc.(Eng.)*
NORTHERN IRELAND CENTRE . . *G. H. Moir, J.P.*
SCOTTISH CENTRE . . . . . *R. H. Dean, B.Sc.Tech.*
NORTH SCOTLAND SUB-CENTRE . . . . *P. Philip*
SOUTH MIDLAND CENTRE . . . . . *W. E. Clark*
RUGBY SUB-CENTRE . . . . . . *H. Orchard*
SOUTHERN CENTRE . . . . . . *G. D. Arden*
WESTERN CENTRE (BRISTOL) . . . *A. H. McQueen*
WESTERN CENTRE (CARDIFF) . . . . *D. J. Thomas*
WEST WALES (SWANSEA) SUB-CENTRE . *O. J. Mayo*
SOUTH-WESTERN SUB-CENTRE . . *W. E. Johnson*

Published by The Institution, Savoy Place, London, W.C.2. *Telephone:* Temple Bar 7676. *Telegrams:* "Voltampere, Phone, London."
Printed by Unwin Brothers Limited, Woking and London.